GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:20:48 ; Search time 307.385 Seconds
(without alignments)
10509.347 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	111	100.0	169620	2	AC012674	AC012674 Homo sapi
c	2	48	43.2	143372	9	AL137847	AL137847 Human DNA
c	3	40	36.0	123779	30	AC021025	AC021025 Homo sapi
	4	40	36.0	128118	2	AC076969	AC076969 Homo sapi
c	5	40	36.0	128583	9	AC121249	AC121249 Homo sapi
c	6	40	36.0	148290	9	AC107979	AC107979 Homo sapi
c	7	40	36.0	165649	9	AC103996	AC103996 Homo sapi
	8	40	36.0	178650	9	AC104303	AC104303 Homo sapi
c	9	40	36.0	192826	9	AC090762	AC090762 Homo sapi
	10	39	35.1	32918	2	AC007445	AC007445 Homo sapi
	11	39	35.1	38936	9	AL358817	AL358817 Human DNA
	12	39	35.1	124271	2	AC025179	AC025179 Homo sapi
	13	39	35.1	146671	9	AC008814	AC008814 Homo sapi
c	14	39	35.1	159747	2	AP001019	AP001019 Homo sapi
	15	39	35.1	162740	2	AC034249	AC034249 Homo sapi
c	16	39	35.1	169772	9	AC069538	AC069538 Homo sapi
c	17	39	35.1	175466	9	AL607077	AL607077 Human DNA
c	18	38	34.2	110000	2	AL691517_2	Continuation (3 of
	19	38	34.2	166706	9	AC068875	AC068875 Homo sapi
	20	38	34.2	207408	2	AC068618	AC068618 Homo sapi
c	21	38	34.2	207548	9	AC087283	AC087283 Homo sapi
c	22	36	32.4	121720	9	AL591491	AL591491 Human DNA
c	23	35	31.5	153940	9	AC022294	AC022294 Homo sapi
	24	35	31.5	325069	2	AC079737	AC079737 Homo sapi
	25	34	30.6	99577	9	AC026324	AC026324 Homo sapi
	26	34	30.6	108040	2	AC068150	AC068150 Homo sapi
c	27	34	30.6	134760	9	AC099484	AC099484 Homo sapi
c	28	34	30.6	146059	2	AC019030	AC019030 Homo sapi
	29	34	30.6	172206	9	AC092119	AC092119 Homo sapi
c	30	34	30.6	172567	2	AC015493	AC015493 Homo sapi
	31	34	30.6	173166	9	AC092375	AC092375 Homo sapi
c	32	34	30.6	273807	2	AC025421	AC025421 Homo sapi
	33	34	30.6	316296	2	AC092285	AC092285 Homo sapi
c	34	33	29.7	33458	9	HSDJ60101	AL109656 Human DNA
c	35	33	29.7	49616	9	AL365267	AL365267 Human DNA
c	36	33	29.7	77743	9	HSTCRBV	U03115 Human V bet
	37	33	29.7	153788	2	AC023971	AC023971 Homo sapi
c	38	33	29.7	166434	9	AC020717	AC020717 Homo sapi
c	39	33	29.7	215422	9	U66060	U66060 Human germl
c	40	31	27.9	70313	2	AC016216	AC016216 Homo sapi
	41	31	27.9	82225	9	HS253J14	Z80771 Human DNA s
c	42	31	27.9	164550	9	CNS01RHY	AL162633 Human chr
c	43	31	27.9	173706	2	AC022694	AC022694 Homo sapi
	44	31	27.9	176418	9	AC090811	AC090811 Homo sapi
c	45	31	27.9	184759	9	AC022695	AC022695 Homo sapi

ALIGNMENTS

RESULT 1
AC012674/c
LOCUS AC012674 169620 bp DNA linear HTG 07-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP1-458H3, WORKING DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AC012674
VERSION AC012674.10 GI:9719580
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169620)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,

Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169620)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Aug 7, 2000 this sequence version replaced gi:8705345.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: HMQG
Center clone name: RP1-458H3
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 139025 bases at least Q40
Consensus quality: 154842 bases at least Q30
Consensus quality: 159725 bases at least Q20
Estimated insert size: 162720; sum-of-contigs estimation
Estimated insert size: 171608; agarose-fp estimation
Quality coverage: 3.9x in Q20 bases; agarose-fp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 28689: contig of 28689 bp in length
* 28690 28789: gap of unknown length
* 28790 50832: contig of 22043 bp in length
* 50833 50932: gap of unknown length
* 50933 69144: contig of 18212 bp in length
* 69145 69244: gap of unknown length
* 69245 84204: contig of 14960 bp in length
* 84205 84304: gap of unknown length
* 84305 94667: contig of 10363 bp in length
* 94668 94767: gap of unknown length
* 94768 107261: contig of 12494 bp in length
* 107262 107361: gap of unknown length
* 107362 117550: contig of 10189 bp in length
* 117551 117650: gap of unknown length
* 117651 126939: contig of 9289 bp in length
* 126940 127039: gap of unknown length
* 127040 135040: contig of 8001 bp in length
* 135041 135140: gap of unknown length
* 135141 141639: contig of 6499 bp in length
* 141640 141739: gap of unknown length
* 141740 149558: contig of 7819 bp in length
* 149559 149658: gap of unknown length
* 149659 154562: contig of 4904 bp in length
* 154563 154662: gap of unknown length
* 154663 158987: contig of 4325 bp in length
* 158988 159087: gap of unknown length
* 159088 162376: contig of 3289 bp in length
* 162377 162476: gap of unknown length
* 162477 165191: contig of 2715 bp in length
* 165192 165291: gap of unknown length
* 165292 167173: contig of 1882 bp in length
* 167174 167273: gap of unknown length
* 167274 168393: contig of 1120 bp in length
* 168394 168493: gap of unknown length
* 168494 169620: contig of 1127 bp in length.

FEATURES Location/Qualifiers
source 1. .169620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP1-458H3"

BASE COUNT 52024 a 33180 c 32128 g 50322 t 1966 others
ORIGIN

Query Match 100.0%; Score 111; DB 2; Length 169620;
Best Local Similarity 100.0%; Pred. No. 2.3e-53;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87441 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 87382

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87381 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 87331
```

RESULT 2
AL137847/c
LOCUS AL137847 143372 bp DNA linear PRI 16-NOV-2001
DEFINITION Human DNA sequence from clone RP11-439K3 on chromosome 9q22.2-31.1, complete sequence.
ACCESSION AL137847
VERSION AL137847.12 GI:16973786
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143372)
AUTHORS Kimberley,A.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 17, 2001 this sequence version replaced gi:16408610.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP

```
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-439K3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-439K3 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-439K3 is at 1 in this sequence. The
            true left end of clone RP11-344L17 is at 141373 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .143372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-439K3"
                     /clone_lib="RPCI-11.2"
BASE COUNT    42809 a  30224 c  28152 g  42187 t
ORIGIN

  Query Match             43.2%;  Score 48;  DB 9;  Length 143372;
  Best Local Similarity   100.0%;  Pred. No. 9.6e-17;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   82306 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 82259


RESULT 3
AC021025/c
ID   AC021025   standard; DNA; HTG; 123779 BP.
XX
AC   AC021025;
XX
SV   AC021025.9
XX
DT   14-JAN-2000 (Rel. 62, Created)
DT   08-AUG-2000 (Rel. 64, Last updated, Version 11)
XX
DE   Homo sapiens chromosome 3 clone RP11-79K17, WORKING DRAFT SEQUENCE, 10
DE   unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-123779
RA   Muzny D.M., Adams C., Bailey M., Barbaria J., Blankenburg K., Bodota B.,
RA   Bouck J., Bowie S., Brooks A., Buhay C., Bunac C., Burkett C., Burrows J.,
RA   Carter M., Chacko J., Chen Z., Cox C., David R., Delgado O., Deshazo D.,
RA   Ding Y., Domah-Rashid N., Dugan-Rocha S., Durbin K.J., Fernandez C.,
RA   Ferraguto D., Forcum-Tansey J., Frantz P., Ganesh R., Gorrell J.H.,
RA   Gorrell L.L., Guevara W., Harris K., Hernandez J., Hodgson A., Hogues M.,
RA   Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M.,
RA   Kelly S., Kondejewski N., Kong Y., Kovar C., Leal B., Li Z., Lichtarge O.,
RA   Liu J., Liu W., Logan O., Lozado R.J., Lu J., Lucier R., Martin R.,
RA   Martinez C., McLeod M.P., Mei G., Morgan M., Morris S., Nash S., Nelson A.,
RA   Nguyen R., Nguyen N., Nguyen S., Oswal G., Parish B., Paxton S., Payton B.,
RA   Perez L., Pu L.L., Quiles M., Reiter D., Rives M., Samuel S., Say J.,
RA   Scherer S., Shah E., Shen H., Simon M., Sparks A., Stamps A., Sucgang R.,
RA   Tabor P., Taylor T., Vasquez L., Vinson R., Vo Q., Wahbah M.,
RA   Watlington S., Weinstock G., Weinstock I.R., Williamson A., Worley K.,
RA   Wren J., Wrensford G., Yu W., Zhou X., Nelson D., Gibbs R.;
RT   "Direct Submission";
RL   Unpublished.
XX
RN   [2]
RP   1-123779
RA   Worley K.C.;
RT   ;
RL   Submitted (12-JAN-2000) to the EMBL/GenBank/DDBJ databases.
RL   Human Genome Sequencing Center, Department of Molecular and Human Genetics,
RL   Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
XX
CC   On Aug 7, 2000 this sequence version replaced gi:8700003.
CC   -------------- Genome Center
CC   Center: Baylor College of Medicine
CC   Center code: BCM
CC   Web site: http://www.hgsc.bcm.tmc.edu/
CC   Contact: hgsc-help@bcm.tmc.edu
CC   -------------- Project Information
CC   Center project name: HMXB
CC   Center clone name: RP11-79K17
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; L08821
CC   Chemistry: Dye-terminator Big Dye: 100% of reads
CC   Assembly program: Phrap; version 0.990329
CC   Consensus quality: 112332 bases at least Q40
CC   Consensus quality: 117272 bases at least Q30
CC   Consensus quality: 119789 bases at least Q20
CC   Estimated insert size: 121963; sum-of-contigs estimation
CC   Quality coverage: 0x in Q20 bases; agarose-fp estimation
CC   Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation
CC   --------------
CC   * NOTE: Estimated insert size may differ from sequence length
CC   *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 10 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1    35931: contig of 35931 bp in length
CC   *    35932    36031: gap of unknown length
CC   *    36032    60754: contig of 24723 bp in length
CC   *    60755    60854: gap of unknown length
CC   *    60855    71667: contig of 10813 bp in length
CC   *    71668    71767: gap of unknown length
CC   *    71768    84826: contig of 13059 bp in length
CC   *    84827    84926: gap of unknown length
CC   *    84927    96721: contig of 11795 bp in length
CC   *    96722    96821: gap of unknown length
CC   *    96822   105594: contig of 8773 bp in length
CC   *   105595   105694: gap of unknown length
CC   *   105695   112907: contig of 7213 bp in length
CC   *   112908   113007: gap of unknown length
CC   *   113008   120198: contig of 7191 bp in length
CC   *   120199   120298: gap of unknown length
CC   *   120299   122220: contig of 1922 bp in length
CC   *   122221   122320: gap of unknown length
CC   *   122321   123779: contig of 1459 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .123779
FT                   /chromosome="3"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-79K17"
XX
SQ   Sequence 123779 BP; 37802 A; 24017 C; 23437 G; 37612 T; 911 other;

  Query Match             36.0%;  Score 40;  DB 30;  Length 123779;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
```

```
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db    3083 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 3044


RESULT 4
AC076969
LOCUS       AC076969              128118 bp    DNA     linear   HTG 15-OCT-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K12, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC076969
VERSION     AC076969.6  GI:16117967
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128118)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128118)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:10047573.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBRJ
                Center clone name: RP11-79K12
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 111307 bases at least Q40
                Consensus quality: 117905 bases at least Q30
                Consensus quality: 121494 bases at least Q20
                Estimated insert size: 122854; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20354: contig of 20354 bp in length
            *    20355    20454: gap of unknown length
            *    20455    33071: contig of 12617 bp in length
            *    33072    33171: gap of unknown length
            *    33172    46935: contig of 13764 bp in length
            *    46936    47035: gap of unknown length
            *    47036    57066: contig of 10031 bp in length
            *    57067    57166: gap of unknown length
            *    57167    70378: contig of 13212 bp in length
            *    70379    70478: gap of unknown length
            *    70479    81786: contig of 11308 bp in length
            *    81787    81886: gap of unknown length
            *    81887    91428: contig of 9542 bp in length
            *    91429    91528: gap of unknown length
            *    91529   100455: contig of 8927 bp in length
            *   100456   100555: gap of unknown length
            *   100556   107879: contig of 7324 bp in length
            *   107880   107979: gap of unknown length
            *   107980   114149: contig of 6170 bp in length
            *   114150   114249: gap of unknown length
            *   114250   120186: contig of 5937 bp in length
            *   120187   120286: gap of unknown length
            *   120287   123236: contig of 2950 bp in length
            *   123237   123336: gap of unknown length
            *   123337   125606: contig of 2270 bp in length
            *   125607   125706: gap of unknown length
            *   125707   128118: contig of 2412 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128118
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K12"
BASE COUNT    35433 a  25053 c  23792 g  42367 t   1473 others
ORIGIN

  Query Match             36.0%;  Score 40;  DB 2;  Length 128118;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db   55106 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 55145


RESULT 5
AC121249/c
LOCUS       AC121249              128583 bp    DNA     linear   PRI 01-JUN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K17, complete sequence.
```

ACCESSION AC121249 AC021025
VERSION AC121249.1 GI:20806313
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 128583)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128583)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 128583)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On May 16, 2002 this sequence version replaced gi:9719675.
---------- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
Drafting Center: BCM
---------- Project Information
Center project name: chr-3
Center clone name: RP11-79K17 (bc0196)
---------- Summary Statistics
Sequencing vector: plasmid; 45% of reads
Sequencing vector: M13; L08821; 55% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 128548 bases at least Q40
Consensus quality: 128580 bases at least Q30
Consensus quality: 128581 bases at least Q20
Insert size: 128583; sum-of-contigs
Quality coverage: 13.4x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-147N17 (UWGC:bc0267) AC104300, 22116-bp overlap
3': RP11-391P4 (UWGC:bc0402) AC104303, 114502-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
12049	12136	4077	4013	8696	8902
579	<800	2067	2145	6	<800
1056	1065	9698	9932	1323	1337
13085	13126	3729	3734	1024	961
10944	10836	1632	1616	8421	8408
581	<800	8069	8139	1114	1126
882	887	1507	1485	3538	3534
8435	8467	11939	11871	6051	6022
3924	4103	133	<800	5129	5048
9842	9759	2135	2145	58	<800
4401	4377	78	<800	1895	1903
4458	4377	5873	5861	799	<800
4587	4522	896	902	858	885
6048	5947	3312	3356	931	885
1395	1385	1505	1485	859	885
3059	3071	6694	6716	3934	3936
3911	3861	10032	9932	1065	1058
138	<800	2168	2145	2629	2723
50	<800	1794	1785	859	885
582	<800	227	<800	6855	6888
2225	2229	4002	4013	7821	7819
984	967	2714	2726	7210	7201
887	887	5628	5624	6992	6888
1042	1065	743	758	1204	1224
2783	2808	3045	2963	2722	2723
7074	7158	1288	1262	321	<800
4087	3861	967	973	1066	1058
7976	7926	461	<800	470	<800
1233	1199	4649	4618	3055	3072
3875	3861	5289	5417	10961	10848
27	<800	1098	1095	6725	6888

1137	1065	798	809	1896	1903
750	749	111	<800	8682	8666
214	<800	7644	7609	4438	4342
836	887	10084	9932	514	<800
1670	1651	3641	3589	3077	3072
2650	2609	7555	7609	1371	1337
1890	1872			4185	4087
4615	4522			719	<800
1321	1273			4130	4087
				367	<800
				2979	3072
				333	<800

```
FEATURES             Location/Qualifiers
     source          1. .128583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    37777 a  24752 c  24789 g  41265 t
ORIGIN

  Query Match             36.0%;  Score 40;  DB 9;  Length 128583;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db   27283 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 27244


RESULT 6
AC107979/c
LOCUS       AC107979               148290 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone CTD-3049M7, complete sequence.
ACCESSION   AC107979
VERSION     AC107979.7  GI:21592043
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-3049M7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell;P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 26, 2002 this sequence version replaced gi:21321840.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24533
                Center clone name: 3049_M_7
            --------------
```

```
FEATURES             Location/Qualifiers
     source          1. .148290
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-3049M7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   complement(1982. .2092)
                     /rpt_family="MIR"
     repeat_region   complement(3356. .3654)
                     /rpt_family="AluSx"
     repeat_region   3789. .3844
                     /rpt_family="MER58"
     repeat_region   complement(4429. .4728)
                     /rpt_family="AluSp"
     repeat_region   5494. .5723
                     /rpt_family="(TA)n"
     repeat_region   5727. .5766
                     /rpt_family="(GA)n"
     repeat_region   5806. .5857
                     /rpt_family="AT_rich"
     repeat_region   6377. .6581
                     /rpt_family="L1"
     repeat_region   6647. .7069
                     /rpt_family="L1ME3"
     repeat_region   7057. .7631
                     /rpt_family="L1ME3"
     repeat_region   complement(8821. .8958)
                     /rpt_family="MIR3"
     repeat_region   11129. .11224
                     /rpt_family="FRAM"
     repeat_region   11229. .11338
                     /rpt_family="CT-rich"
     repeat_region   11943. .12090
                     /rpt_family="MIR"
     repeat_region   12818. .13060
                     /rpt_family="MIR"
     repeat_region   13118. .19146
                     /rpt_family="L1PA2"
     repeat_region   20846. .20912
                     /rpt_family="CT-rich"
     repeat_region   21294. .21314
                     /rpt_family="AT_rich"
     repeat_region   21635. .23417
                     /rpt_family="L1PA16"
     repeat_region   23438. .25940
                     /rpt_family="L1PA16"
     repeat_region   26121. .26145
                     /rpt_family="(TTCA)n"
     repeat_region   26293. .26417
                     /rpt_family="MIR"
     repeat_region   complement(26872. .26937)
                     /rpt_family="L2"
     repeat_region   complement(27064. .27431)
                     /rpt_family="L2"
     repeat_region   27917. .28357
                     /rpt_family="MLT1B"
     repeat_region   28460. .28806
                     /rpt_family="MLT1I"
     repeat_region   29509. .29698
                     /rpt_family="MIR"
     repeat_region   29716. .29790
                     /rpt_family="CT-rich"
     repeat_region   29791. .29811
                     /rpt_family="AT_rich"
     repeat_region   complement(31182. .31787)
                     /rpt_family="L2"
     repeat_region   complement(32398. .32581)
                     /rpt_family="AluSq"
     repeat_region   complement(32582. .32777)
                     /rpt_family="L1PB1"
     repeat_region   complement(32778. .32912)
                     /rpt_family="AluSq"
     repeat_region   33104. .33207
                     /rpt_family="L1MC5"
     repeat_region   complement(33208. .33365)
                     /rpt_family="MER53"
     repeat_region   33366. .33529
                     /rpt_family="L1MC5"
     repeat_region   33877...34506
                     /rpt_family="L1MEd"
     repeat_region   34667. .34808
                     /rpt_family="L1MEd"
     repeat_region   35585. .36698
                     /rpt_family="L1M4"
     repeat_region   complement(36699. .36830)
                     /rpt_family="FLAM_C"
     repeat_region   36831. .38141
                     /rpt_family="L1M4"
     repeat_region   38194. .38373
                     /rpt_family="MLT1I"
     repeat_region   39530. .39757
                     /rpt_family="MIR"
     repeat_region   39992. .40017
                     /rpt_family="(TTTTG)n"
     repeat_region   40283. .40583
                     /rpt_family="AluY"
     repeat_region   41414. .41616
                     /rpt_family="MLT1C"

  Query Match             36.0%;  Score 40;  DB 9;  Length 148290;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           ||||||||||||||||||||||||||||||||||||||||
Db   84499 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 84460


RESULT 7
AC103996/c
LOCUS       AC103996              165649 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-76E17, complete sequence.
ACCESSION   AC103996
VERSION     AC103996.7  GI:21637504
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-76E17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165649)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
```

```
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 1, 2002 this sequence version replaced gi:21592191.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21917
                Center clone name: 76_E_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .165649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-76E17"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(2. .865)
                     /rpt_family="L1PA13"
     repeat_region   complement(864. .894)
                     /rpt_family="L1PA13"
     repeat_region   complement(895. .1248)
                     /rpt_family="THE1A"
     repeat_region   complement(1249. .1504)
                     /rpt_family="L1PA13"
     repeat_region   1505. .1626
                     /rpt_family="AluSx"
     repeat_region   1627. .1659
                     /rpt_family="(CAAA)n"
     repeat_region   1660. .1831
                     /rpt_family="AluSx"
     repeat_region   complement(1832. .3132)
                     /rpt_family="L1PA13"
     repeat_region   complement(3140. .3699)
                     /rpt_family="L1M4c"
     repeat_region   complement(4461. .4755)
                     /rpt_family="AluSx"
     repeat_region   5027. .5397
                     /rpt_family="L2"
     repeat_region   complement(5650. .6330)
                     /rpt_family="L1ME1"
     repeat_region   complement(6343. .6460)
                     /rpt_family="L1ME1"
     repeat_region   7214. .7319
                     /rpt_family="L1MC3"
     repeat_region   7354. .7655
                     /rpt_family="L1MC3"
     repeat_region   7659. .7743
                     /rpt_family="L1PA10"
     repeat_region   complement(7745. .7843)
                     /rpt_family="AluSp/q"
     repeat_region   7844. .14156
                     /rpt_family="L1PA10"
     repeat_region   14198. .14328
                     /rpt_family="AluY"
     repeat_region   14340. .14503
                     /rpt_family="(TA)n"
     repeat_region   14507. .15032
                     /rpt_family="L1MC3"
     repeat_region   15036. .15220
                     /rpt_family="L1MC3"
     repeat_region   complement(15218. .16449)
                     /rpt_family="L1PA4"
     repeat_region   16450. .18154
                     /rpt_family="L1PA4"
     repeat_region   18155. .19210
                     /rpt_family="L1MC3"
     repeat_region   19211. .19266
                     /rpt_family="(TA)n"
     repeat_region   19267. .19372
                     /rpt_family="L1MC3"
     repeat_region   19401. .19476
                     /rpt_family="(TTATA)n"
     repeat_region   19483. .19546
                     /rpt_family="(CATATA)n"
     repeat_region   19596. .19625
                     /rpt_family="AT_rich"
     repeat_region   19696. .19750
                     /rpt_family="GA-rich"
     repeat_region   19752. .19920
                     /rpt_family="L1MD3"
     repeat_region   19994. .20102
                     /rpt_family="L2"
     unsure          20049. .20114
                     /note="single clone coverage"
     repeat_region   20485. .20655
                     /rpt_family="MIR3"
     repeat_region   complement(21285. .21441)
                     /rpt_family="MIR"
     repeat_region   21496. .21717
                     /rpt_family="L2"
     repeat_region   complement(21720. .21796)
                     /rpt_family="MIR"
```

```
     repeat_region   complement(21943. .22131)
                     /rpt_family="MIR"
     repeat_region   23082. .23195
                     /rpt_family="L2"
     repeat_region   23198. .23248
                     /rpt_family="GA-rich"
     repeat_region   23267. .23333
                     /rpt_family="(CATA)n"
     repeat_region   complement(24465. .24833)
                     /rpt_family="MLT1A2"
     repeat_region   26142. .26334
                     /rpt_family="MIR"
     repeat_region   complement(27659. .27811)
                     /rpt_family="MIR"
     repeat_region   complement(28811. .28880)
                     /rpt_family="MIR"
     repeat_region   complement(29026. .29269)
                     /rpt_family="MIR"
     repeat_region   complement(29673. .29706)
                     /rpt_family="MSTB"

  Query Match             36.0%;  Score 40;  DB 9;  Length 165649;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           |||||||||||||||||||||||||||||||||||||||
Db    8110 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 8071


RESULT 8
AC104303
LOCUS       AC104303               178650 bp    DNA     linear   PRI 25-FEB-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-391P4, complete sequence.
ACCESSION   AC104303 AC064830
VERSION     AC104303.2  GI:18874945
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Feb 25, 2002 this sequence version replaced gi:17402782.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: WUGSC
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-391P4 (bc0402)
            -------------- Summary Statistics
                Sequencing vector: unknown; 55% of reads
                Sequencing vector: plasmid; 45% of reads
                Chemistry: Dye-terminator ET; 89% of reads
                Chemistry: Dye-terminator Big Dye; 11% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 178494 bases at least Q40
                Consensus quality: 178631 bases at least Q30
                Consensus quality: 178650 bases at least Q20
                Insert size: 178648; sum-of-contigs
                Quality coverage: 8.0x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  RP11-475O23 (UWGC:bc0439) AC023346
                3':  RP11-79K12 AC076969
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
7050	7417	8696	8715	7163	7098
2067	2138	6	<800	6382	6518
13472	13264	3077	3083	512	<800
7644	7859	514	<800	449	<800
111	<800	4438	4377	7988	7884
798	783	8682	8715	11779	11698
1098	1109	1896	1876	472	<800
5288	5395	6724	6948	783	787
4649	4507	10961	10773	926	926
461	<800	3055	3083	2431	2538
967	982	470	<800	1015	1029
1288	1268	1066	1037	758	<800
3045	2923	321	<800	1550	1541
743	783	2722	2683	361	<800

5628	5682	1204	1300	763	<800
2714	2699	6992	7281	2975	3002
4002	3900	7210	7830	3299	3468
227	<800	7821	8101	1907	1894
1794	1784	6855	6948	4020	4220
2168	2138	859	857	5158	5126
10032	9744	2629	2683	497	<800
6694	6810	1065	1037	4321	4220
1505	1491	3934	3965	887	926
3312	3304	859	857	180	<800
896	912	931	935	4969	4887
5873	5938	858	857	4755	4674
78	<800	799	857	26	<800
2135	2138	1895	1876	906	926
133	<800	58	<800	2325	2346
11939	11575	5129	5076	369	<800
1507	1491	6051	6048	2523	2538
8069	8181	3538	3537	1889	1894
1632	1617	1114	1094	4437	4428
3729	3678	8420	8391	373	<800
5553	5395	1024	1037	24	<800
358	<800	1323	1380	263	<800
3764	3678	1502	1481	632	<800
7022	6810	4538	4539	1258	1247
8534	8821	1414	1380	1114	1083
5206	5050	521	<800	1370	1320
9052	9744	8346	8391	1262	1247
1625	1617	39	<800	1546	1541
749	783	2716	2683	459	<800
10190	9744	2546	2683	135	<800
3857	3795	1064	1037	3142	3183
6144	6490	926	935	109	<800
514	<800	5154	5076	7858	7884
1364	1342	1975	1981	1891	1894
669	<800	628	<800	723	<800
		2398	2399	6975	7098
		8148	8391	1878	1894
		1713	1673	3747	4012
		135	<800	3146	3183
		566	<800	3450	3737
		1115	1094	390	<800
		2344	2399	560	<800
		1119	1188	7249	7098

Query Match 36.0%; Score 40; DB 9; Length 178650;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db 165449 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 165488
```

RESULT 9
AC090762/c
LOCUS AC090762 192826 bp DNA linear PRI 28-FEB-2002
DEFINITION Homo sapiens chromosome 15, clone RP11-387E8, complete sequence.
ACCESSION AC090762
VERSION AC090762.9 GI:18997378
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192826)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-387E8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 192826)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 192826)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,

```
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2002 this sequence version replaced gi:18377189.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12392
                Center clone name: 387_E_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192826
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-387E8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(864. .1222)
                     /rpt_family="L1PB3"
     repeat_region   1223. .1252
                     /rpt_family="(TAGG)n"
     repeat_region   complement(1253. .1492)
                     /rpt_family="L1PB3"
     repeat_region   4287. .4576
                     /rpt_family="AluJb"
     repeat_region   4671. .5034
                     /rpt_family="THE1B"
     repeat_region   5130. .5438
                     /rpt_family="AluSx"
     repeat_region   5517. .5662
                     /rpt_family="MIR"
     repeat_region   5798. .5827
                     /rpt_family="AT_rich"
     repeat_region   5985. .6114
                     /rpt_family="AluJb"
     repeat_region   6227. .6253
                     /rpt_family="(CA)n"
     repeat_region   7329. .7446
                     /rpt_family="MIR"
     repeat_region   complement(8147. .8452)
                     /rpt_family="AluJo"
     repeat_region   8540. .8639
                     /rpt_family="MER45"
     repeat_region   complement(8832. .8916)
                     /rpt_family="L2"
     repeat_region   10391. .10599
                     /rpt_family="MER3"
     repeat_region   11688. .11828
                     /rpt_family="L1MC/D"
     repeat_region   12029. .12078
                     /rpt_family="AT_rich"
     repeat_region   12092. .12447
                     /rpt_family="THE1C"
     repeat_region   complement(13616. .13751)
                     /rpt_family="MIR3"
     repeat_region   complement(13958. .14142)
                     /rpt_family="MIR"
     repeat_region   complement(15277. .15553)
                     /rpt_family="MER8"
     repeat_region   complement(15811. .16005)
                     /rpt_family="MIR"
     repeat_region   complement(16101. .16440)
                     /rpt_family="L3"
     repeat_region   16920. .16958
                     /rpt_family="(TCCC)n"
     repeat_region   complement(17145. .17444)
                     /rpt_family="AluSx"
     repeat_region   complement(18418. .19953)
                     /rpt_family="L1MEc"
     repeat_region   complement(19978. .20262)
                     /rpt_family="L1MEc"
     repeat_region   complement(20288. .20794)
                     /rpt_family="L1MEc"
     repeat_region   complement(20822. .21097)
                     /rpt_family="L1MEc"
     repeat_region   complement(21345. .21743)
                     /rpt_family="L1MEc"
     repeat_region   22599. .23518
                     /rpt_family="L1MEc"
     repeat_region   23527. .23901
                     /rpt_family="L1MEc"
     repeat_region   complement(23927. .24026)
                     /rpt_family="MSTB1"
     repeat_region   complement(24027. .24256)
                     /rpt_family="MER30"
     repeat_region   complement(24257. .24557)
                     /rpt_family="MSTB1"
     repeat_region   24563. .24594
                     /rpt_family="AT_rich"
     repeat_region   complement(25258. .25639)
                     /rpt_family="L1MCc"
     repeat_region   complement(25688. .25838)
                     /rpt_family="L1MCc"
     repeat_region   26167. .26506
                     /rpt_family="THE1B"
     repeat_region   26746. .27094
                     /rpt_family="Tigger2a"
     repeat_region   27095. .27184
                     /rpt_family="MADE1"
     repeat_region   27185. .27279
                     /rpt_family="Tigger2a"
     repeat_region   28321. .28363
                     /rpt_family="(TG)n"
     repeat_region   29172. .29333
                     /rpt_family="MIR"
     repeat_region   30664. .30898
                     /rpt_family="L1MB8"
     repeat_region   31577. .31598
                     /rpt_family="AT_rich"
     repeat_region   32378. .32475
                     /rpt_family="CT-rich"
     repeat_region   34655. .34960
                     /rpt_family="AluY"
     repeat_region   complement(35296. .35389)
                     /rpt_family="MER5B"
     repeat_region   complement(35419. .35517)
                     /rpt_family="L1MC4"
     repeat_region   complement(35518. .35823)
                     /rpt_family="AluSx"
     repeat_region   complement(35824. .36030)
                     /rpt_family="L1MC4"
     repeat_region   36682. .36758
                     /rpt_family="MER5A"
     repeat_region   36975. .37160
                     /rpt_family="MER5A"
     repeat_region   38140. .38353
```

```
                     /rpt_family="MER96B"
     repeat_region   38633. .38656
                     /rpt_family="(TTTA)n"
     repeat_region   complement(38657. .40056)
                     /rpt_family="L1PA5"
     repeat_region   complement(40060. .40250)

  Query Match             36.0%;  Score 40;  DB 9;  Length 192826;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           ||||||||||||||||||||||||||||||||||||||||
Db 190946 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 190907


RESULT 10
AC007445
LOCUS       AC007445               32918 bp    DNA     linear   HTG 30-JUN-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-344B7 map 18, *** SEQUENCING
            IN PROGRESS ***, 1 ordered piece.
ACCESSION   AC007445
VERSION     AC007445.5  GI:8844149
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-344B7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
    TLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8705092.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L571
                Center clone name: 344_B_7
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *          1    32918: contig of 32918 bp in length.
FEATURES                Location/Qualifiers
     source             1. .32918
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="18"
                        /map="18"
                        /clone="RP11-344B7"
                        /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     9462 a   6493 c   6865 g   9853 t    245 others
ORIGIN

  Query Match             35.1%;  Score 39;  DB 2;  Length 32918;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db    3535 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 3573


RESULT 11
AL358817
LOCUS       AL358817               38936 bp    DNA     linear   PRI 06-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-399N22 on chromosome 10,
            complete sequence.
ACCESSION   AL358817
VERSION     AL358817.18  GI:15990637
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38936)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2001 this sequence version replaced gi:14669268.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-399N22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-399N22 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-432J9 is at 36937 in this sequence.
            The true right end of clone RP11-91A1 is at 2000 in this sequence.
FEATURES                Location/Qualifiers
     source             1. .38936
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-399N22"
                     /clone_lib="RPCI-11.2"
BASE COUNT     9315 a    9079 c    9111 g   11431 t
ORIGIN

  Query Match             35.1%;  Score 39;  DB 9;  Length 38936;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db   28896 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 28934


RESULT 12
AC025179
LOCUS       AC025179               124271 bp    DNA     linear   HTG 20-APR-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2174B5, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   AC025179
VERSION     AC025179.4  GI:13699647
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124271)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124271)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 20, 2001 this sequence version replaced gi:7711794.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 694394
            Center clone name: CITB-H1_2174B5
            -------
            Summary Statistics
            Consensus quality: 116609 bases at least Q40
            Consensus quality: 120226 bases at least Q30
            Consensus quality: 121274 bases at least Q20
            Estimated insert size: 117160; agarose-fp estimation
            Estimated insert size: 123571; sum-of-contigs estimation
            Quality coverage:  6.46 in Q20 bases; agarose-fp estimation
            Quality coverage:  6.12 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1052: contig of 1052 bp in length
            *     1053     1152: gap of unknown length
            *     1153     3331: contig of 2179 bp in length
            *     3332     3431: gap of unknown length
            *     3432     7370: contig of 3939 bp in length
            *     7371     7470: gap of unknown length
            *     7471    15483: contig of 8013 bp in length
            *    15484    15583: gap of unknown length
            *    15584    24916: contig of 9333 bp in length
            *    24917    25016: gap of unknown length
            *    25017    39922: contig of 14906 bp in length
            *    39923    40022: gap of unknown length
            *    40023    68684: contig of 28662 bp in length
            *    68685    68784: gap of unknown length
            *    68785   124271: contig of 55487 bp in length.
FEATURES             Location/Qualifiers
     source          1. .124271
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2174B5"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    36863 a   24233 c   23599 g   38876 t    700 others
ORIGIN

  Query Match             35.1%;  Score 39;  DB 2;  Length 124271;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
           |||||||||||||||||||||||||||||||||||||||
Db   79646 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 79684


RESULT 13
AC008814
LOCUS       AC008814               146671 bp    DNA     linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2117L12, complete sequence.
ACCESSION   AC008814
VERSION     AC008814.6  GI:16554342
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146671)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146671)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 146671)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 31, 2001 this sequence version replaced gi:15290309.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.3% of Sequence;
            Estimated Total Number of Errors is 0.7.
FEATURES             Location/Qualifiers
     source          1. .146671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2117L12"
BASE COUNT    44951 a   28815 c   28434 g   44471 t
ORIGIN

  Query Match             35.1%;  Score 39;  DB 9;  Length 146671;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
           |||||||||||||||||||||||||||||||||||||||
Db   54480 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 54518
```

```
RESULT 14
AP001019/c
LOCUS       AP001019              159747 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-752I11 map 18p11.3, WORKING
            DRAFT SEQUENCE, 24 unordered pieces.
ACCESSION   AP001019
VERSION     AP001019.2  GI:8117689
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-752I11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159747)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 159,747 genomic DNA of 18p11.3
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 159747)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997769.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-752I11
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 145356 bases at least Q40
                Consensus quality: 152227 bases at least Q30
                Consensus quality: 155843 bases at least Q20
                Insert size: 157447; sum-of-contigs
                Quality coverage: 4.51x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            24 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     18444 contig of  18444 bp in length
              18545     29631 contig of  11087 bp in length
              29732     43143 contig of  13412 bp in length
              43244     57170 contig of  13927 bp in length
              57271     68240 contig of  10970 bp in length
              68341     78134 contig of   9794 bp in length
              78235     87873 contig of   9639 bp in length
              87974     96197 contig of   8224 bp in length
              96298    103976 contig of   7679 bp in length
             104077    110726 contig of   6650 bp in length
             110827    116866 contig of   6040 bp in length
             116967    122789 contig of   5823 bp in length
             122890    127665 contig of   4776 bp in length
             127766    132618 contig of   4853 bp in length
             132719    136051 contig of   3333 bp in length
             136152    139749 contig of   3598 bp in length
             139850    142810 contig of   2961 bp in length
             142911    145855 contig of   2945 bp in length
             145956    148840 contig of   2885 bp in length
             148941    151243 contig of   2303 bp in length
             151344    153454 contig of   2111 bp in length
             153555    156134 contig of   2580 bp in length
             156235    158089 contig of   1855 bp in length
             158190    159747 contig of   1558 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    18444: contig of 18444 bp in length
            *   18445 18544: gap of      100 bp
            *   18545    29631: contig of 11087 bp in length
            *   29632 29731: gap of      100 bp
            *   29732    43143: contig of 13412 bp in length
            *   43144 43243: gap of      100 bp
            *   43244    57170: contig of 13927 bp in length
            *   57171 57270: gap of      100 bp
            *   57271    68240: contig of 10970 bp in length
            *   68241 68340: gap of      100 bp
            *   68341    78134: contig of 9794 bp in length
            *   78135 78234: gap of      100 bp
            *   78235    87873: contig of 9639 bp in length
            *   87874 87973: gap of      100 bp
            *   87974    96197: contig of 8224 bp in length
            *   96198 96297: gap of      100 bp
            *   96298   103976: contig of 7679 bp in length
            *  103977 104076: gap of      100 bp
            *  104077   110726: contig of 6650 bp in length
            *  110727 110826: gap of      100 bp
            *  110827   116866: contig of 6040 bp in length
            *  116867 116966: gap of      100 bp
            *  116967   122789: contig of 5823 bp in length
            *  122790 122889: gap of      100 bp
            *  122890   127665: contig of 4776 bp in length
            *  127666 127765: gap of      100 bp
            *  127766   132618: contig of 4853 bp in length
            *  132619 132718: gap of      100 bp
            *  132719   136051: contig of 3333 bp in length
            *  136052 136151: gap of      100 bp
            *  136152   139749: contig of 3598 bp in length
            *  139750 139849: gap of      100 bp
            *  139850   142810: contig of 2961 bp in length
            *  142811 142910: gap of      100 bp
            *  142911   145855: contig of 2945 bp in length
            *  145856 145955: gap of      100 bp
            *  145956   148840: contig of 2885 bp in length
            *  148841 148940: gap of      100 bp
            *  148941   151243: contig of 2303 bp in length
            *  151244 151343: gap of      100 bp
            *  151344   153454: contig of 2111 bp in length
            *  153455 153554: gap of      100 bp
            *  153555   156134: contig of 2580 bp in length
            *  156135 156234: gap of      100 bp
            *  156235   158089: contig of 1855 bp in length
            *  158090 158189: gap of      100 bp
            *  158190   159747: contig of 1558 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159747
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18p11.3"
                     /clone="RP11-752I11"
     misc_feature    1. .18444
                     /note="assembly_fragment"
     misc_feature    18545. .29631
                     /note="assembly_fragment"
```

```
     misc_feature    29732. .43143
                     /note="assembly_fragment"
     misc_feature    43244. .57170
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    57271. .68240
                     /note="assembly_fragment"
     misc_feature    68341. .78134
                     /note="assembly_fragment"
     misc_feature    78235. .87873
                     /note="assembly_fragment"
     misc_feature    87974. .96197
                     /note="assembly_fragment"
     misc_feature    96298. .103976
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    104077. .110726
                     /note="assembly_fragment"
     misc_feature    110827. .116866
                     /note="assembly_fragment"
     misc_feature    116967. .122789
                     /note="assembly_fragment"
     misc_feature    122890. .127665
                     /note="assembly_fragment"
     misc_feature    127766. .132618
                     /note="assembly_fragment"
     misc_feature    132719. .136051
                     /note="assembly_fragment"
     misc_feature    136152. .139749
                     /note="assembly_fragment"
     misc_feature    139850. .142810
                     /note="assembly_fragment"
     misc_feature    142911. .145855
                     /note="assembly_fragment"
     misc_feature    145956. .148840
                     /note="assembly_fragment"
     misc_feature    148941. .151243
                     /note="assembly_fragment"
     misc_feature    151344. .153454
                     /note="assembly_fragment"
     misc_feature    153555. .156134
                     /note="assembly_fragment"
     misc_feature    156235. .158089
                     /note="assembly_fragment"
     misc_feature    158190. .159747
                     /note="assembly_fragment"
BASE COUNT    47302 a  30397 c  29897 g  49849 t   2302 others
ORIGIN

  Query Match            35.1%;  Score 39;  DB 2;  Length 159747;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db  74932 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 74894


RESULT 15
AC034249
LOCUS       AC034249              162740 bp    DNA     linear   HTG 31-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-427C17, WORKING DRAFT
            SEQUENCE, 6 ordered pieces.
ACCESSION   AC034249
VERSION     AC034249.3  GI:15383785
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162740)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162740)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 31, 2001 this sequence version replaced gi:9211234.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 570398
            Center clone name: RPCI-11_427C17
            -------
            Summary Statistics
            Consensus quality: 158733 bases at least Q40
            Consensus quality: 161469 bases at least Q30
            Consensus quality: 162087 bases at least Q20
            Estimated insert size: 160000; pulse field gel estimation
            Estimated insert size: 162240; sum-of-contigs estimation
            Quality coverage:  9.28 in Q20 bases; pulse field gel estimation
            Quality coverage:  9.16 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    85871: contig of 85871 bp in length
            *    85872    85971: gap of unknown length
            *    85972    96765: contig of 10794 bp in length
            *    96766    96865: gap of unknown length
            *    96866   103302: contig of 6437 bp in length
            *   103303   103402: gap of unknown length
            *   103403   111971: contig of 8569 bp in length
            *   111972   112071: gap of unknown length
            *   112072   129034: contig of 16963 bp in length
            *   129035   129134: gap of unknown length
            *   129135   162740: contig of 33606 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-427C17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    48722 a  31435 c  31563 g  50520 t    500 others
ORIGIN

  Query Match            35.1%;  Score 39;  DB 2;  Length 162740;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-11;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
           |||||||||||||||||||||||||||||||||||||||
Db   3599 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 3637


Search completed: April 25, 2003, 00:30:53
Job time : 719.885 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:19:38 ; Search time 51.0118 Seconds
(without alignments)
4900.271 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	100.0	447	21	AAC03794	Human secreted pro
	2	111	100.0	447	21	AAZ42680	Human 5' EST isola
c	3	24	21.6	660	23	AAS73441	DNA encoding novel
	4	20	18.0	2526	23	AAS66674	DNA encoding novel
	5	18	16.2	431	24	ABK24383	DNA encoding human
	6	18	16.2	431	24	ABK24390	DNA encoding human
c	7	18	16.2	570	22	ABA63453	Human foetal liver
c	8	18	16.2	570	22	ABA30652	Probe #9118 for ge
c	9	18	16.2	570	22	AAK11985	Human brain expres
c	10	18	16.2	570	22	AAK37688	Human bone marrow
c	11	18	16.2	570	22	AAI18447	Probe #8380 for ge
c	12	18	16.2	570	22	AAI43563	Probe #12249 used
c	13	18	16.2	570	24	ABS11680	Human genome-deriv
	14	18	16.2	827	23	AAS87115	DNA encoding novel
c	15	18	16.2	3342	23	AAS87118	DNA encoding novel
	16	17	15.3	266	20	AAV89709	EST clone CT857.
c	17	17	15.3	531	23	ABK43012	Genomic sequence #
c	18	17	15.3	609	24	ABK80793	Bacillus clausii g
	19	17	15.3	828	22	AAI94407	Human neuroblastom
	20	17	15.3	1047	24	AAS62681	cDNA sequence #468
	21	17	15.3	1363	9	AAN70128	Novel DNA encoding
c	22	17	15.3	1546	22	AAK82125	Human immune/haema
c	23	17	15.3	1982	21	AAC68089	Human secreted pro
	24	17	15.3	6928	22	ABA21109	Human nervous syst
	25	17	15.3	21340	23	ABL12924	Drosophila melanog
	26	17	15.3	62909	22	AAF28545	Genomic fragment #
	27	17	15.3	172637	24	ABN83124	Human voltage-acti
	28	17	15.3	495269	24	ABQ67195	Listeria innocua c
	29	17	15.3	1503900	22	AAK95240	Human neuregulin-1
	30	17	15.3	1503900	22	AAK96733	Human neuregulin-1
c	31	17	15.3	3011208	24	ABQ69245	Listeria innocua D
	32	16	14.4	79	22	AAC89232	Human brain T calc
c	33	16	14.4	250	16	AAT22213	Human gene signatu
c	34	16	14.4	260	24	ABN18202	Human ORFX polynuc
	35	16	14.4	281	24	ABL67351	Thyroid cancer rel
c	36	16	14.4	302	24	ABN78873	Human ORF3820 cDNA
c	37	16	14.4	347	24	ABN60208	Human cancer relat
c	38	16	14.4	348	22	AAF65225	Novel human polynu
c	39	16	14.4	360	22	AAK57925	Human immune/haema
	40	16	14.4	377	24	ABL38444	Human colon tumour
	41	16	14.4	407	23	ABV36584	Human prostate exp
	42	16	14.4	414	22	AAI82407	Human polynucleoti
c	43	16	14.4	415	23	ABV07896	Human prostate exp
c	44	16	14.4	432	23	ABV36813	Human prostate exp
c	45	16	14.4	434	22	AAI79941	Human polynucleoti

ALIGNMENTS

RESULT 1
AAC03794
ID AAC03794 standard; cDNA; 447 BP.
XX
AC AAC03794;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3792.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG03788.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3792; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 111; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 3e-46;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 2
AAZ42680
ID AAZ42680 standard; cDNA; 447 BP.
XX
AC AAZ42680;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:439.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
OS Homo sapiens.
XX
PN WO9953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
DR P-PSDB; AAY65066.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 402; 837pp; English.
XX
CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 111; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 3e-46;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 3
AAS73441/c
ID AAS73441 standard; cDNA; 660 BP.
XX
AC AAS73441;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9245.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG09254.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 9245; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 660 BP; 149 A; 127 C; 148 G; 236 T; 0 other;

Query Match 21.6%; Score 24; DB 23; Length 660;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 AAAACTCCTGGGTCTCTGTGTGTG 92
||||||||||||||||||||||||
Db 656 AAAACTCCTGGGTCTCTGTGTGTG 633

RESULT 4
AAS66674
ID AAS66674 standard; cDNA; 2526 BP.
XX
AC AAS66674;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #2478.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
WO200175067-A2.

11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG02487.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 2478; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2526 BP; 871 A; 536 C; 541 G; 578 T; 0 other;

Query Match 18.0%; Score 20; DB 23; Length 2526;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 TTTCATCTTTGCAGGGACTT 51
||||||||||||||||||||
Db 2377 TTTCATCTTTGCAGGGACTT 2396

RESULT 5
ABK24383
ID ABK24383 standard; cDNA; 431 BP.
XX
AC ABK24383;
XX
DT 09-APR-2002 (first entry)
XX
DE DNA encoding human lung cancer protein, Seq ID No 80.
XX
KW Human; lung cancer; cytostatic; vaccine; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192525-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US17066.
XX
PR 26-MAY-2000; 2000US-207485P.
PR 06-SEP-2000; 2000US-230475P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Harlocker SL, Wang T, Bangur CS, Klee JI, Switzer A;
XX
DR WPI; 2002-122068/16.
XX
PT New lung tumour polypeptides and polynucleotides, useful in
PT pharmaceutical compositions, such as vaccines, for treating or
PT preventing lung cancer, or as probes or primers for nucleic acid
PT hybridisation -
XX
PS Claim 9; Page 172; 179pp; English.
XX
CC The invention relates to novel human lung cancer polynucleotide (I)
CC and polypeptides (II). (I) and (II) are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis and treatment of lung
CC cancer. The polynucleotides are also useful as probes or primers for
CC nucleic acid hybridisation. ABK24314-ABK24397 represent human lung
CC cancer coding sequences of the invention.

```
XX
SQ   Sequence 431 BP; 108 A; 97 C; 123 G; 101 T; 2 other;

  Query Match            16.2%;  Score 18;  DB 24;  Length 431;
  Best Local Similarity  100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 6
ABK24390
ID   ABK24390 standard; cDNA; 431 BP.
XX
AC   ABK24390;
XX
DT   09-APR-2002  (first entry)

     DNA encoding human lung cancer protein, Seq ID No 87.
XX
KW   Human; lung cancer; cytostatic; vaccine; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192525-A2.
XX
PD   06-DEC-2001.
XX
PF   25-MAY-2001; 2001WO-US17066.
XX
PR   26-MAY-2000; 2000US-207485P.
PR   06-SEP-2000; 2000US-230475P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Harlocker SL,  Wang T,  Bangur CS,  Klee JI,  Switzer A;
XX
DR   WPI; 2002-122068/16.
XX
PT   New lung tumour polypeptides and polynucleotides, useful in
PT   pharmaceutical compositions, such as vaccines, for treating or
PT   preventing lung cancer, or as probes or primers for nucleic acid
PT   hybridisation -
XX
PS   Claim 1; Page 174; 179pp; English.

     The invention relates to novel human lung cancer polynucleotide (I)
     and polypeptides (II). (I) and (II) are useful in pharmaceutical
CC   compositions, such as vaccines, for the diagnosis and treatment of lung
CC   cancer. The polynucleotides are also useful as probes or primers for
CC   nucleic acid hybridisation. ABK24314-ABK24397 represent human lung
CC   cancer coding sequences of the invention.
XX
SQ   Sequence 431 BP; 108 A; 97 C; 123 G; 101 T; 2 other;

  Query Match            16.2%;  Score 18;  DB 24;  Length 431;
  Best Local Similarity  100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 7
ABA63453/c
ID   ABA63453 standard; DNA; 570 BP.
XX
AC   ABA63453;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #11758.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 1; SEQ ID NO 11758; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match            16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity  100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 8
ABA30652/c
ID   ABA30652 standard; DNA; 570 BP.
XX
AC   ABA30652;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #9118 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
```

```
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 1; SEQ ID No 9118; 530pp; English.

     The present invention relates to single exon nucleic acid probes for
     measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 9
AAK11985/c
ID   AAK11985 standard; DNA; 570 BP.

     AAK11985;

DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 11976.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 11976; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 10
AAK37688/c
ID   AAK37688 standard; DNA; 570 BP.
XX
AC   AAK37688;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 12245.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 12245; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
```

```
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 11
AAI18447/c
ID   AAI18447 standard; DNA; 570 BP.
XX
AC   AAI18447;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #8380 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.

     Penn SG,  Hanzel DK,  Chen W,  Rank DR;

DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 8380; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 12
AAI43563/c
ID   AAI43563 standard; DNA; 570 BP.
XX
AC   AAI43563;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #12249 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 12249; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             16.2%;  Score 18;  DB 22;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 13
ABS11680/c
ID   ABS11680 standard; DNA; 570 BP.
XX
AC   ABS11680;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human genome-derived single exon probe from lung SEQ ID No 11671.
XX
```

KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 1; SEQ ID No 11671; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
the array; identifying exons in a eukaryotic genome, comprising
(a) algorithmically predicting at least one exon from genomic sequences
of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Query Match 16.2%; Score 18; DB 24; Length 570;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 TGTGTGTGCCTGAGTGGC 102
||||||||||||||||||
Db 505 TGTGTGTGCCTGAGTGGC 488

RESULT 14
AAS87115
ID AAS87115 standard; cDNA; 827 BP.
XX
AC AAS87115;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22919.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22928.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 22919; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 827 BP; 198 A; 202 C; 249 G; 178 T; 0 other;

Query Match 16.2%; Score 18; DB 23; Length 827;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 TGCCTGAGTGGCTGCTCT 108
||||||||||||||||||
Db 344 TGCCTGAGTGGCTGCTCT 361

RESULT 15
AAS87118/c
ID AAS87118 standard; cDNA; 3342 BP.
XX
AC AAS87118;
XX
[redacted] 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #22922.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG22931.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
[redacted] responsible for genetic disorders or other traits and to assess
biodiversity -
XX
PS Claim 1; SEQ ID No 22922; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3342 BP; 898 A; 796 C; 867 G; 781 T; 0 other;

Query Match 16.2%; Score 18; DB 23; Length 3342;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 TGCCTGAGTGGCTGCTCT 108
||||||||||||||||||
Db 113 TGCCTGAGTGGCTGCTCT 96

Search completed: April 25, 2003, 00:00:24
Job time : 56.0118 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:49:04 ; Search time 373.284 Seconds
(without alignments)
4815.901 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

[redacted]d size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	100.0	470	17	AQ770688	AQ770688 HS_5368_B
	2	45	40.5	525	17	AQ165256	AQ165256 HS_3025_B
	3	38	34.2	628	17	AQ237815	AQ237815 RPCI11-70
	4	35	31.5	410	17	AQ442274	AQ442274 HS_5137_A
	5	35	31.5	453	17	AQ437684	AQ437684 HS_5137_A
c	6	35	31.5	635	17	AQ390599	AQ390599 CITBI-E1-
c	7	34	30.6	529	17	AQ881246	AQ881246 HS_5137_B
c	8	33	29.7	482	17	AQ320567	AQ320567 RPCI11-99
c	9	33	29.7	563	17	AQ420187	AQ420187 RPCI-11-1
c	10	33	29.7	723	17	AQ386439	AQ386439 RPCI11-15
	11	31	27.9	553	17	AZ521751	AZ521751 RPCI-11-1
	12	31	27.9	695	17	AG179297	AG179297 Pan trogl
	13	30	27.0	360	17	AQ207172	AQ207172 HS_3239_B
c	14	30	27.0	399	17	AQ115544	AQ115544 RPCI11-57
c	15	30	27.0	435	17	AQ116061	AQ116061 RPCI11-57
	16	30	27.0	551	17	AQ569689	AQ569689 HS_5333_B
	17	29	26.1	414	17	AQ003326	AQ003326 RPCI11-25
c	18	29	26.1	615	17	AG161224	AG161224 Pan trogl
c	19	28	25.2	363	17	AQ120796	AQ120796 HS_3076_A
	20	28	25.2	376	17	AQ548294	AQ548294 RPCI-11-4
	21	28	25.2	401	17	AQ588089	AQ588089 CITBI-E1-
	22	28	25.2	419	17	AQ436209	AQ436209 HS_5049_B
c	23	28	25.2	530	17	AQ193128	AQ193128 HS_3060_B
	24	28	25.2	653	17	AG160919	AG160919 Pan trogl
	25	27	24.3	553	17	AQ238365	AQ238365 RPCI11-63
	26	27	24.3	606	17	AQ350708	AQ350708 RPCI11-11
	27	26	23.4	478	17	B75615	B75615 RPCI11-11L1
c	28	26	23.4	653	17	AG143347	AG143347 Pan trogl
	29	26	23.4	690	17	AG111263	AG111263 Pan trogl
	30	25	22.5	411	17	AQ715895	AQ715895 HS_5430_B
	31	25	22.5	427	17	AQ697116	AQ697116 HS_5528_A
	32	25	22.5	456	17	AQ707169	AQ707169 HS_5564_A
	33	25	22.5	468	17	AQ533287	AQ533287 RPCI-11-3
	34	25	22.5	519	17	AQ519549	AQ519549 HS_5168_A
	35	25	22.5	520	17	AQ375479	AQ375479 RPCI-11-1
	36	25	22.5	552	17	AQ378145	AQ378145 RPCI11-15
	37	25	22.5	563	17	AQ379371	AQ379371 RPCI11-16
	38	25	22.5	583	17	AG070759	AG070759 Pan trogl
	39	25	22.5	609	17	AG155327	AG155327 Pan trogl
	40	25	22.5	747	17	AG170366	AG170366 Pan trogl
	41	24	21.6	413	17	AQ266658	AQ266658 RPCI11-73
c	42	24	21.6	425	17	AQ683450	AQ683450 HS_5432_B
	43	24	21.6	471	17	AQ677950	AQ677950 HS_5531_B
	44	24	21.6	506	17	AQ284282	AQ284282 RPCI11-79
	45	24	21.6	541	17	AQ686074	AQ686074 HS_5537_B

ALIGNMENTS

RESULT 1
AQ770688
LOCUS AQ770688 470 bp DNA linear GSS 28-JUL-1999
DEFINITION HS_5368_B2_C08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=944 Col=16 Row=F, DNA sequence.
ACCESSION AQ770688
VERSION AQ770688.1 GI:5648804
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 944  row: F  column: 16
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=944 Col=16 Row=F"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT       83 a    112 c    131 g    141 t      3 others
ORIGIN

  Query Match             100.0%;  Score 111;  DB 17;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-47;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 162

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 213


RESULT 2
AQ165256
LOCUS       AQ165256                525 bp    DNA     linear   GSS 16-OCT-1998
DEFINITION  HS_3025_B2_G06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=12 Row=N, DNA sequence.
ACCESSION   AQ165256
VERSION     AQ165256.1  GI:3563451
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3025  row: N  column: 12
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES              Location/Qualifiers
     source           1. .525
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3025 Col=12 Row=N"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      102 a    139 c    137 g    143 t      4 others
ORIGIN

  Query Match             40.5%;  Score 45;  DB 17;  Length 525;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-13;
  Matches   45;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 TCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 78
           |||||||||||||||||||||||||||||||||||||||||||||
Db      99 TCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 143


RESULT 3
AQ237815
LOCUS       AQ237815                628 bp    DNA     linear   GSS 21-APR-1999
DEFINITION  RPCI11-70H4.TK RPCI-11 Homo sapiens genomic clone RPCI-11-70H4, DNA
            sequence.
ACCESSION   AQ237815
VERSION     AQ237815.1  GI:3670106
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .628
                      /organism="Homo sapiens"
                      /db_xref="GDB:7526667"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-70H4"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      125 a    150 c    164 g    189 t
ORIGIN

  Query Match             34.2%;  Score 38;  DB 17;  Length 628;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-09;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           ||||||||||||||||||||||||||||||||||||||
Db     157 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 194


RESULT 4
AQ442274
LOCUS       AQ442274                410 bp    DNA     linear   GSS 31-MAR-1999
```

```
DEFINITION  HS_5137_A1_F12_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=713 Col=23 Row=K, DNA sequence.
ACCESSION   AQ442274
VERSION     AQ442274.1  GI:4553613
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 713  row: K  column: 23
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=713 Col=23 Row=K"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       80 a    111 c    111 g    107 t      1 others
ORIGIN

  Query Match             31.5%;  Score 35;  DB 17;  Length 410;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           |||||||||||||||||||||||||||||||||||
Db     160 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 194


RESULT 5
AQ437684
LOCUS       AQ437684                 453 bp    DNA     linear   GSS 31-MAR-1999
DEFINITION  HS_5137_A2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=713 Col=12 Row=O, DNA sequence.
ACCESSION   AQ437684
VERSION     AQ437684.1  GI:4549023
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 713  row: O  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=713 Col=12 Row=O"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       84 a    127 c    117 g    124 t      1 others
ORIGIN

  Query Match             31.5%;  Score 35;  DB 17;  Length 453;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           |||||||||||||||||||||||||||||||||||
Db     163 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 197


RESULT 6
AQ390599/c
LOCUS       AQ390599                 635 bp    DNA     linear   GSS 06-MAR-1999
DEFINITION  CITBI-E1-2544B15.TR CITBI-E1 Homo sapiens genomic clone 2544B15,
            DNA sequence.
ACCESSION   AQ390599
VERSION     AQ390599.1  GI:4361622
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2544B15.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
```

```
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .635
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="2544B15"
                      /clone_lib="CITBI-E1"
                      /sex="male"
                      /cell_type="sperm"
                      /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                      CalTech Human BAC Library D"
BASE COUNT      192 a    144 c    172 g    127 t
ORIGIN

  Query Match             31.5%;  Score 35;  DB 17;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-07;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           |||||||||||||||||||||||||||||||||||
Db     527 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 493


RESULT 7
AQ881246/c
LOCUS       AQ881246                529 bp    DNA     linear   GSS 09-NOV-1999
DEFINITION  HS_5137_B1_F08_T7 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=8905 Col=15 Row=L, DNA sequence.
ACCESSION   AQ881246
VERSION     AQ881246.1  GI:6312713
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 8905  row: L  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 529.
FEATURES              Location/Qualifiers
     source           1. .529
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=8905 Col=15 Row=L"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      137 a    147 c    121 g    117 t      7 others
ORIGIN

  Query Match             30.6%;  Score 34;  DB 17;  Length 529;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-07;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGC 93
           |||||||||||||||||||||||||||||||||
Db     382 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGC 349


RESULT 8
AQ320567/c
LOCUS       AQ320567                482 bp    DNA     linear   GSS 04-MAY-1999
DEFINITION  RPCI11-99N1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-99N1, DNA
            sequence.
ACCESSION   AQ320567
VERSION     AQ320567.1  GI:4050696
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .482
                      /organism="Homo sapiens"
                      /db_xref="GDB:7537944"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-99N1"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      142 a    121 c    106 g    112 t      1 others
ORIGIN

  Query Match             29.7%;  Score 33;  DB 17;  Length 482;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
Db     338 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 306


RESULT 9
AQ420187/c
LOCUS       AQ420187                563 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-185J19.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-185J19
```

```
          , DNA sequence.
ACCESSION   AQ420187
VERSION     AQ420187.1  GI:4477911
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-185J19.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .563
                      /organism="Homo sapiens"
                      /db_xref="GDB:7570890"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-185J19"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      173 a    127 c    114 g    149 t
ORIGIN

  Query Match             29.7%;  Score 33;  DB 17;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
Db     333 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 301


RESULT 10
AQ386439/c
LOCUS       AQ386439                723 bp    DNA     linear   GSS 21-MAY-1999
DEFINITION  RPCI11-154D6.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-154D6,
            DNA sequence.
ACCESSION   AQ386439
VERSION     AQ386439.1  GI:4357462
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-154D6.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .723
                      /organism="Homo sapiens"
                      /db_xref="GDB:7558829"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-154D6"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      224 a    165 c    135 g    199 t
ORIGIN

  Query Match             29.7%;  Score 33;  DB 17;  Length 723;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-06;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
Db     331 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 299


RESULT 11
AZ521751
LOCUS       AZ521751                553 bp    DNA     linear   GSS 16-OCT-2000
DEFINITION  RPCI-11-175G22.TVB RPCI-11 Homo sapiens genomic clone
            RPCI-11-175G22, DNA sequence.
ACCESSION   AZ521751
VERSION     AZ521751.1  GI:10834261
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     BAC end sequences of library RPCI-11
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-175G22.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            This BAC end was generated during the R&D process and may have
            higher chance of clone tracking errors.
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
```

```
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="GDB:7566981"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-175G22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      109 a    148 c    137 g    158 t      1 others
ORIGIN

  Query Match             27.9%;  Score 31;  DB 17;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-05;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           ||||||||||||||||||||||||||||||
       140 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 170


RESULT 12
AG179297
LOCUS       AG179297                695 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-051I11.TJ, genomic survey
            sequence.
ACCESSION   AG179297
VERSION     AG179297.1  GI:16708977
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-051I11.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 695)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .695
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-051I11.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      243 a    135 c    122 g    192 t      3 others
ORIGIN

  Query Match             27.9%;  Score 31;  DB 17;  Length 695;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-05;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGT 91
           |||||||||||||||||||||||||||||||
Db     171 GAGTATGTAAAACTCCTGGGTCTCTGTGTGT 201


RESULT 13
AQ207172
LOCUS       AQ207172                360 bp    DNA     linear   GSS 18-SEP-1998
DEFINITION  HS_3239_B1_C03_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3239 Col=5 Row=F, DNA sequence.
ACCESSION   AQ207172
VERSION     AQ207172.1  GI:3618377
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3239  row: F  column: 5
            Class: BAC ends
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3239 Col=5 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       63 a     93 c    107 g     93 t      4 others
ORIGIN

  Query Match             27.0%;  Score 30;  DB 17;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-05;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 GGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||
Db     171 GGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 200


RESULT 14
AQ115544/c
LOCUS       AQ115544                399 bp    DNA     linear   GSS 20-APR-1999
DEFINITION  RPCI11-57K21.TK RPCI-11 Homo sapiens genomic clone RPCI-11-57K21,
            DNA sequence.
ACCESSION   AQ115544
VERSION     AQ115544.1  GI:3491665
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
```

```
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-57K21.TJ
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Homo sapiens"
                     /db_xref="GDB:7521764"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-57K21"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      120 a    111 c     95 g     73 t
ORIGIN

  Query Match             27.0%;  Score 30;  DB 17;  Length 399;
  Best Local Similarity   100.0%;  Pred. No. 5e-05;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 90
           ||||||||||||||||||||||||||||||
Db     380 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 351


RESULT 15
AQ116061/c
LOCUS       AQ116061                435 bp    DNA     linear   GSS 20-APR-1999
DEFINITION  RPCI11-57L19.TK RPCI-11 Homo sapiens genomic clone RPCI-11-57L19,
            DNA sequence.
ACCESSION   AQ116061
VERSION     AQ116061.1  GI:3492182
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-57L19.TJ
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /db_xref="GDB:7521786"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-57L19"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      131 a    115 c     103 g     86 t
ORIGIN

  Query Match             27.0%;  Score 30;  DB 17;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 5e-05;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 90
           ||||||||||||||||||||||||||||||
Db     379 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 350


Search completed: April 25, 2003, 00:52:48
Job time : 378.284 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:57:29 ; Search time 13.1361 Seconds
(without alignments)
2591.419 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	16	14.4	536	4	US-09-221-017B-1081	Sequence 1081, Ap
	2	16	14.4	997	4	US-09-149-476-307	Sequence 307, App
	3	16	14.4	2128	3	US-09-233-506-1	Sequence 1, Appli
	4	16	14.4	2178	4	US-09-334-601-6	Sequence 6, Appli
c	5	16	14.4	2288	4	US-09-334-601-1	Sequence 1, Appli
	6	16	14.4	2915	1	US-07-746-705A-16	Sequence 16, Appl
	7	16	14.4	2915	2	US-08-380-182-18	Sequence 18, Appl
c	8	16	14.4	3494	4	US-09-334-601-5	Sequence 5, Appli
c	9	16	14.4	5977	3	US-09-024-020B-1	Sequence 1, Appli
c	10	16	14.4	5977	4	US-09-425-043-1	Sequence 1, Appli
c	11	16	14.4	6007	3	US-09-024-020B-2	Sequence 2, Appli
c	12	16	14.4	6007	4	US-09-425-043-2	Sequence 2, Appli
c	13	16	14.4	6556	3	US-09-024-020B-7	Sequence 7, Appli
c	14	16	14.4	6556	4	US-09-425-043-7	Sequence 7, Appli
c	15	16	14.4	6586	3	US-09-024-020B-43	Sequence 43, Appl
c	16	16	14.4	6586	4	US-09-425-043-43	Sequence 43, Appl
	17	16	14.4	6822	4	US-09-426-998-3	Sequence 3, Appli
c	18	16	14.4	6826	3	US-09-024-020B-8	Sequence 8, Appli
c	19	16	14.4	6826	4	US-09-425-043-8	Sequence 8, Appli
	20	16	14.4	7741	4	US-09-426-998-4	Sequence 4, Appli
	21	16	14.4	12847	1	US-08-550-715-1	Sequence 1, Appli
c	22	16	14.4	15567	4	US-09-627-376-3	Sequence 3, Appli
c	23	15	13.5	441	4	US-09-222-575-109	Sequence 109, App
c	24	15	13.5	1112	1	US-08-136-743B-1	Sequence 1, Appli
c	25	15	13.5	1335	2	US-08-743-637B-182	Sequence 182, App
c	26	15	13.5	1725	2	US-08-907-166-11	Sequence 11, Appl
	27	15	13.5	1751	2	US-08-955-138-3	Sequence 3, Appli
c	28	15	13.5	2031	3	US-08-262-220-13	Sequence 13, Appl
c	29	15	13.5	2031	3	US-08-471-733-13	Sequence 13, Appl
c	30	15	13.5	2031	3	US-08-468-878-13	Sequence 13, Appl
c	31	15	13.5	2031	4	US-08-750-494-13	Sequence 13, Appl
c	32	15	13.5	2075	3	US-08-262-220-3	Sequence 3, Appli
c	33	15	13.5	2075	3	US-08-471-733-3	Sequence 3, Appli
c	34	15	13.5	2075	3	US-08-468-878-3	Sequence 3, Appli
c	35	15	13.5	2075	4	US-08-750-494-3	Sequence 3, Appli
c	36	15	13.5	3213	4	US-09-442-100-5	Sequence 5, Appli
	37	15	13.5	4403	2	US-08-284-941-1	Sequence 1, Appli
	38	15	13.5	4403	2	US-08-447-642-1	Sequence 1, Appli
	39	15	13.5	4403	4	US-09-236-503-1	Sequence 1, Appli
	40	15	13.5	4403	5	PCT-US93-02147A-1	Sequence 1, Appli
c	41	15	13.5	6350	2	US-08-385-335A-9	Sequence 9, Appli
c	42	15	13.5	9636	1	US-08-323-170B-1	Sequence 1, Appli
c	43	15	13.5	9636	4	US-08-954-441-1	Sequence 1, Appli
	44	15	13.5	10409	3	US-08-772-440-33	Sequence 33, Appl
	45	15	13.5	28001	4	US-09-819-993-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-221-017B-1081/c
; Sequence 1081, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1081:
;    SEQUENCE CHARACTERISTICS:
;      .LENGTH:  536 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...536
US-09-221-017B-1081

Query Match 14.4%; Score 16; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 ATTCTTTTTCATCTTT 41
||||||||||||||||
Db 287 ATTCTTTTTCATCTTT 272

ULT 2
09-149-476-307
Sequence 307, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: PZ002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911

```
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,636
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,874
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,910
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,864
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,631
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,845
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,892
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,761
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/047,595
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,599
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,588
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,585
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,586
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,590
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,594
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,589
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,593
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/047,614
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/043,578
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/043,576
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/047,501
;  EARLIER FILING DATE: 1997-05-23
;  EARLIER APPLICATION NUMBER: 60/043,670
;  EARLIER FILING DATE: 1997-04-11
;  EARLIER APPLICATION NUMBER: 60/056,632
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,664
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,876
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,881
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,909
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,875
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,862
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,887
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/056,908
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/057,650
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/056,884
;  EARLIER FILING DATE: 1997-08-22
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match            14.4%;  Score 16;  DB 4;  Length 997;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 GATCTTTTGCCTTGCA 23
            ||||||||||||||||
Db     662 GATCTTTTGCCTTGCA 677


RESULT 3
US-09-233-506-1
; Sequence 1, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
;  APPLICANT: Fukuda, Minoru
;  APPLICANT: Yeh, Jiunn-Chern
;  TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminytransferase That Forms
;  TITLE OF INVENTION: Core 2, Core 4 and I Branches
;  FILE REFERENCE: P-LJ 3415
;  CURRENT APPLICATION NUMBER: US/09/233,506
;  CURRENT FILING DATE: 1999-01-19
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2128
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (354)..(1670)
US-09-233-506-1

  Query Match            14.4%;  Score 16;  DB 3;  Length 2128;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 GATCTTTTGCCTTGCA 23
            ||||||||||||||||
Db    1821 GATCTTTTGCCTTGCA 1836


RESULT 4
US-09-334-601-6/c
; Sequence 6, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
;  APPLICANT: Kapitonov, Dmitri
;  APPLICANT: Yu, Robert
;  TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;  FILE REFERENCE: VCUIP-6
;  CURRENT APPLICATION NUMBER: US/09/334,601
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 2178
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-334-601-6

  Query Match            14.4%;  Score 16;  DB 4;  Length 2178;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTT 41
            ||||||||||||||||
Db    2150 ATTCTTTTTCATCTTT 2135


RESULT 5
```

```
US-09-334-601-1/c
; Sequence 1, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
;  APPLICANT: Kapitonov, Dmitri
;  APPLICANT: Yu, Robert
;  TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;  FILE REFERENCE: VCUIP-6
;  CURRENT APPLICATION NUMBER: US/09/334,601
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2288
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (29)..(1282)
;  09-334-601-1

  Query Match             14.4%;  Score 16;  DB 4;  Length 2288;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTT 41
           ||||||||||||||||
Db    2260 ATTCTTTTTCATCTTT 2245


RESULT 6
US-07-746-705A-16
; Sequence 16, Application US/07746705A
; Patent No. 5451516
;  GENERAL INFORMATION:
;    APPLICANT:  Matthews, Benjamin F.
;    APPLICANT:  Weisemann, Jane M.
;    TITLE OF INVENTION:  A Recombinant DNA Molecule Encoding
;    TITLE OF INVENTION:  a Bifunctional Plant Enzyme: Aspartokinase and Homoserine
;    TITLE OF INVENTION:  Dehydrogenase
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janelle S. Graeter
;      STREET:  Bldg. 005. Roon 402, BARC-W
;      CITY:  Beltsville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20705
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/746,705A
;      FILING DATE:  19910816
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Graeter, Janelle S.
;      REGISTRATION NUMBER:  35,024
;      REFERENCE/DOCKET NUMBER:  4000.91
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301)504-5676
;      TELEFAX:  (301)504-5060
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2915 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..2593
US-07-746-705A-16

  Query Match             14.4%;  Score 16;  DB 1;  Length 2915;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 CTTTGCAGGGACTTCT 53
           ||||||||||||||||
Db    1977 CTTTGCAGGGACTTCT 1992


RESULT 7
US-08-380-182-18
; Sequence 18, Application US/08380182
; Patent No. 5858749
;  GENERAL INFORMATION:
;    APPLICANT:  Matthews, Benjamin F.
;    APPLICANT:  Weisemann, Jane M.
;    TITLE OF INVENTION:  A Bifunctional Protein From Carrots
;    TITLE OF INVENTION:  (Daucus carota) with Aspartokinase and Homoserine
;    TITLE OF INVENTION:  Dehydrogenase Activities"
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janelle S. Graeter
;      STREET:  Room 411,BGldg. 005, BARC-W
;      CITY:  Beltsville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20705
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/380,182
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Graeter, Janelle S.
;      REGISTRATION NUMBER:  35,024
;      REFERENCE/DOCKET NUMBER:  0226.94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-504-6629
;      TELEFAX:  301-504-5060
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2915 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Daucus carota
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..2593
US-08-380-182-18

  Query Match             14.4%;  Score 16;  DB 2;  Length 2915;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 CTTTGCAGGGACTTCT 53
           ||||||||||||||||
Db    1977 CTTTGCAGGGACTTCT 1992
```

```
RESULT 8
US-09-334-601-5/c
; Sequence 5, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
;  APPLICANT: Kapitonov, Dmitri
;  APPLICANT: Yu, Robert
;  TITLE OF INVENTION: NOVEL SIALYLTRANSFERASES
;  FILE REFERENCE: VCUIP-6
;  CURRENT APPLICATION NUMBER: US/09/334,601
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 3494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-334-601-5

  Query Match             14.4%;  Score 16;  DB 4;  Length 3494;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTT 41
           ||||||||||||||||
Db    3467 ATTCTTTTTCATCTTT 3452


RESULT 9
US-09-024-020B-1/c
; Sequence 1, Application US/09024020B
; Patent No. 6030810
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/024,020B
;      FILING DATE:  16-FEB-1998
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-024-020B-1

  Query Match             14.4%;  Score 16;  DB 3;  Length 5977;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5814 GTCTCTGTGTGTGCCT 5799


RESULT 10
US-09-425-043-1/c
; Sequence 1, Application US/09425043
; Patent No. 6335172
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/425,043
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/024,020
;      FILING DATE:  16-FEB-1998
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-425-043-1

  Query Match             14.4%;  Score 16;  DB 4;  Length 5977;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5814 GTCTCTGTGTGTGCCT 5799


RESULT 11
US-09-024-020B-2/c
```

```
; Sequence 2, Application US/09024020B
; Patent No. 6030810
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/024,020B
;      FILING DATE:  16-FEB-1998
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6007 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-024-020B-2

  Query Match             14.4%;  Score 16;  DB 3;  Length 6007;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5844 GTCTCTGTGTGTGCCT 5829


RESULT 12
US-09-425-043-2/c
; Sequence 2, Application US/09425043
; Patent No. 6335172
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/425,043
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/024,020
;      FILING DATE:  16-FEB-1998
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6007 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-425-043-2

  Query Match             14.4%;  Score 16;  DB 4;  Length 6007;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5844 GTCTCTGTGTGTGCCT 5829


RESULT 13
US-09-024-020B-7/c
; Sequence 7, Application US/09024020B
; Patent No. 6030810
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/024,020B
;      FILING DATE:  16-FEB-1998
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-024-020B-7

  Query Match             14.4%;  Score 16;  DB 3;  Length 6556;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5961 GTCTCTGTGTGTGCCT 5946


RESULT 14
US-09-425-043-7/c
; Sequence 7, Application US/09425043
; Patent No. 6335172
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/425,043
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/024,020
;      FILING DATE:  16-FEB-1998
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6556 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-425-043-7

  Query Match             14.4%;  Score 16;  DB 4;  Length 6556;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5961 GTCTCTGTGTGTGCCT 5946


RESULT 15
US-09-024-020B-43/c
; Sequence 43, Application US/09024020B
; Patent No. 6030810
;  GENERAL INFORMATION:
;    APPLICANT:  DELGADO, STEPHEN G.
;    APPLICANT:  DIETRICH, PAUL S.
;    APPLICANT:  FISH, LINDA M.
;    APPLICANT:  HERMAN, RONALD C.
;    APPLICANT:  SANGAMESWARAN, LAKSHMI
;    TITLE OF INVENTION:  NOVEL CLONED TETRODOTOXIN-SENSITIVE
;    TITLE OF INVENTION:  SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JANET PAULINE CLARK
;      STREET:  3401 HILLVIEW AVENUE, MS A2-250
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94304-1397
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/024,020B
;      FILING DATE:  16-FEB-1998
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,447
;      FILING DATE:  26-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CLARK, JANET P.
;      REGISTRATION NUMBER:  34,799
;      REFERENCE/DOCKET NUMBER:  R0020B-REG
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 852-3097
;      TELEFAX:  (650) 855-5322
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6586 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-024-020B-43

  Query Match             14.4%;  Score 16;  DB 3;  Length 6586;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCT 95
           ||||||||||||||||
Db    5991 GTCTCTGTGTGTGCCT 5976


Search completed: April 25, 2003, 00:54:06
Job time : 19.1361 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 00:41:49 ; Search time 22.7692 Seconds
(without alignments)
5304.628 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 709820 seqs, 544064369 residues

[redacted] size : 0

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match %	Length	DB	ID	Description
c	1	33	29.7	684973	10	US-09-263-959-1	Sequence 1, Appli
	2	18	16.2	431	10	US-09-866-562-80	Sequence 80, Appl
	3	18	16.2	431	10	US-09-866-562-87	Sequence 87, Appl
c	4	18	16.2	570	10	US-09-864-761-9118	Sequence 9118, Ap
c	5	17	15.3	531	9	US-10-092-154-1899	Sequence 1899, Ap
c	6	17	15.3	531	10	US-09-764-847-1899	Sequence 1899, Ap
c	7	17	15.3	609	10	US-09-974-300-8084	Sequence 8084, Ap
	8	17	15.3	1047	10	US-09-822-830A-468	Sequence 468, App
	9	17	15.3	172637	10	US-09-805-458A-3	Sequence 3, Appli
	10	17	15.3	1503841	9	US-09-946-807-1	Sequence 1, Appli
	11	17	15.3	1503841	10	US-09-795-668-1	Sequence 1, Appli
	12	17	15.3	1503841	10	US-09-795-686-1	Sequence 1, Appli
	13	16	14.4	281	10	US-09-964-824A-385	Sequence 385, App
	14	16	14.4	374	9	US-10-046-935-2033	Sequence 2033, Ap
	15	16	14.4	374	9	US-09-878-178-2033	Sequence 2033, Ap
	16	16	14.4	374	9	US-10-146-502-2033	Sequence 2033, Ap
	17	16	14.4	374	9	US-10-060-036-985	Sequence 985, App
	18	16	14.4	374	9	US-10-060-036-2528	Sequence 2528, Ap
	19	16	14.4	436	9	US-09-918-995-26108	Sequence 26108, A
c	20	16	14.4	447	10	US-09-880-107-871	Sequence 871, App
c	21	16	14.4	481	10	US-09-560-863-45	Sequence 45, Appl
	22	16	14.4	494	9	US-09-918-995-20796	Sequence 20796, A
	23	16	14.4	497	10	US-09-864-761-1356	Sequence 1356, Ap
	24	16	14.4	524	10	US-09-797-207-5	Sequence 5, Appli
	25	16	14.4	548	10	US-09-864-761-12421	Sequence 12421, A
	26	16	14.4	552	10	US-09-998-598-1325	Sequence 1325, Ap
c	27	16	14.4	557	10	US-09-764-877-746	Sequence 746, App
c	28	16	14.4	557	10	US-09-764-877-3495	Sequence 3495, Ap
	29	16	14.4	747	12	US-10-001-879-101	Sequence 101, App
c	30	16	14.4	805	9	US-09-984-245-106	Sequence 106, App
c	31	16	14.4	805	9	US-09-966-262-106	Sequence 106, App
c	32	16	14.4	805	9	US-09-983-966-106	Sequence 106, App
c	33	16	14.4	805	9	US-10-143-090-106	Sequence 106, App
	34	16	14.4	997	9	US-09-809-391-307	Sequence 307, App
	35	16	14.4	1009	10	US-09-764-864-12	Sequence 12, Appl
	36	16	14.4	2002	10	US-09-925-300-592	Sequence 592, App
c	37	16	14.4	2017	9	US-10-102-806-137	Sequence 137, App
	38	16	14.4	2108	10	US-09-797-207-3	Sequence 3, Appli
	39	16	14.4	2147	9	US-09-981-353-43	Sequence 43, Appl
	40	16	14.4	2229	10	US-09-925-297-337	Sequence 337, App
c	41	16	14.4	2268	9	US-10-102-806-142	Sequence 142, App
	42	16	14.4	2287	10	US-09-764-864-473	Sequence 473, App
	43	16	14.4	2319	10	US-09-874-390-1	Sequence 1, Appli
c	44	16	14.4	2340	9	US-09-984-245-51	Sequence 51, Appl
c	45	16	14.4	2340	9	US-09-966-262-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-263-959-1

Query Match 29.7%; Score 33; DB 10; Length 684973;

```
  Best Local Similarity   100.0%;  Pred. No. 2.4e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      66 TGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||
Db 404690 TGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 404658


RESULT 2
US-09-866-562-80
; Sequence 80, Application US/09866562
; Patent No. US20020009758A1
; GENERAL INFORMATION:
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Klee, Jennifer
;  APPLICANT:  Switzer, Anne
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER.
;  FILE REFERENCE: 210121.502
;  CURRENT APPLICATION NUMBER: US/09/866,562
;  CURRENT FILING DATE:  2001-05-25
;  NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 80
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 361,431
;   OTHER INFORMATION: n = A,T,C or G
US-09-866-562-80

  Query Match             16.2%;  Score 18;  DB 10;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 3
US-09-866-562-87
; Sequence 87, Application US/09866562
; Patent No. US20020009758A1
; GENERAL INFORMATION:
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Klee, Jennifer
;  APPLICANT:  Switzer, Anne
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER.
;  FILE REFERENCE: 210121.502
;  CURRENT APPLICATION NUMBER: US/09/866,562
;  CURRENT FILING DATE:  2001-05-25
;  NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 87
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 361,431
;   OTHER INFORMATION: n = A,T,C or G
US-09-866-562-87

  Query Match             16.2%;  Score 18;  DB 10;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 4
US-09-864-761-9118/c
; Sequence 9118, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9118
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000053.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
US-09-864-761-9118
```

```
  Query Match           16.2%;  Score 18;  DB 10;  Length 570;
  Best Local Similarity  100.0%;  Pred. No. 3;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 5
US-10-092-154-1899/c
; Sequence 1899, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1899
;   LENGTH: 531
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1899

  Query Match           15.3%;  Score 17;  DB 9;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      28 TCTTTTTCATCTTTGCA 44
           |||||||||||||||||
Db     141 TCTTTTTCATCTTTGCA 125


RESULT 6
US-09-764-847-1899/c
; Sequence 1899, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1899
;   LENGTH: 531
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1899

  Query Match           15.3%;  Score 17;  DB 10;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      28 TCTTTTTCATCTTTGCA 44
           |||||||||||||||||
Db     141 TCTTTTTCATCTTTGCA 125


RESULT 7
US-09-974-300-8084/c
; Sequence 8084, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8084
;   LENGTH: 609
;   TYPE: DNA
;   ORGANISM: Bacillus clausii
US-09-974-300-8084

  Query Match           15.3%;  Score 17;  DB 10;  Length 609;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      28 TCTTTTTCATCTTTGCA 44
           |||||||||||||||||
Db     536 TCTTTTTCATCTTTGCA 520


RESULT 8
US-09-822-830A-468
; Sequence 468, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 468
;   LENGTH: 1047
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-830A-468

  Query Match           15.3%;  Score 17;  DB 10;  Length 1047;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      76 CTGGGTCTCTGTGTGTG 92
           |||||||||||||||||
Db     828 CTGGGTCTCTGTGTGTG 844


RESULT 9
US-09-805-458A-3
; Sequence 3, Application US/09805458A
; Patent No. US20020042100A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al
;  TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000722
;  CURRENT APPLICATION NUMBER: US/09/805,458A
```

```
;  CURRENT FILING DATE:  2001-03-14
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 172637
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(172637)
;   OTHER INFORMATION: n = A,T,C or G
US-09-805-458A-3

  Query Match             15.3%;  Score 17;  DB 10;  Length 172637;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      19 TTGCAGGATTCTTTTTC 35
           |||||||||||||||||
   150045 TTGCAGGATTCTTTTTC 150061


RESULT 10
US-09-946-807-1
; Sequence 1, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
    OTHER INFORMATION: r=g or a
    FEATURE:
    NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: m=a or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: h=a or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1

  Query Match             15.3%;  Score 17;  DB 9;  Length 1503841;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTTG 42
           |||||||||||||||||
Db 268726 ATTCTTTTTCATCTTTG 268742


RESULT 11
US-09-795-668-1
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/795,668
;  CURRENT FILING DATE:  2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: r=g or a
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: m=a or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
```

```
;   OTHER INFORMATION: h=a or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1

  Query Match             15.3%;  Score 17;  DB 10;  Length 1503841;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTTG 42
           |||||||||||||||||
Db 268726 ATTCTTTTTCATCTTTG 268742


RESULT 12
US-09-795-686-1
; Sequence 1, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2005-001
;  CURRENT APPLICATION NUMBER: US/09/795,686
;  CURRENT FILING DATE:  2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,715
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: r=g or a
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
    OTHER INFORMATION: m=a or c
    NAME/KEY: misc_feature
    LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: h=a or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-795-686-1

  Query Match             15.3%;  Score 17;  DB 10;  Length 1503841;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTTG 42
           |||||||||||||||||
Db 268726 ATTCTTTTTCATCTTTG 268742


RESULT 13
US-09-964-824A-385
; Sequence 385, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
;  APPLICANT: Horrigan, Stephen
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-73
;  CURRENT APPLICATION NUMBER: US/09/964,824A
;  CURRENT FILING DATE:  2001-09-27
;  PRIOR APPLICATION NUMBER: US/60/236,033
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US/60/236,032
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US/60/236,028
;  PRIOR FILING DATE: 2000-09-28
;  NUMBER OF SEQ ID NOS: 583
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 385
;   LENGTH: 281
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-964-824A-385

  Query Match             14.4%;  Score 16;  DB 10;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 ATTCTTTTTCATCTTT 41
           ||||||||||||||||
Db      22 ATTCTTTTTCATCTTT 37


RESULT 14
US-10-046-935-2033
; Sequence 2033, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527C1
;  CURRENT APPLICATION NUMBER: US/10/046,935
;  CURRENT FILING DATE:  2002-01-15
;  NUMBER OF SEQ ID NOS: 2239
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2033
;   LENGTH: 374
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-046-935-2033

  Query Match             14.4%;  Score 16;  DB 9;  Length 374;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 GATCTTTTGCCTTGCA 23
           ||||||||||||||||
```

```
Db    108 GATCTTTTGCCTTGCA 123


RESULT 15
US-09-878-178-2033
; Sequence 2033, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527
;  CURRENT APPLICATION NUMBER: US/09/878,178
;  CURRENT FILING DATE:  2001-06-08
;  NUMBER OF SEQ ID NOS: 2237
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2033
    LENGTH: 374
    TYPE: DNA
    ORGANISM: Homo sapien
US-09-878-178-2033

  Query Match             14.4%;  Score 16;  DB 9;  Length 374;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 GATCTTTTGCCTTGCA 23
          ||||||||||||||||
Db    108 GATCTTTTGCCTTGCA 123


Search completed: April 25, 2003, 02:09:14
Job time : 1043.77 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:20:48 ; Search time 1096.62 Seconds
(without alignments)
10509.347 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_446
Perfect score: 396
Sequence: 1 atgggtggatcttttgcctt.....gamctggatatttcagttga 396

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	299	75.5	169620	2	AC012674	AC012674 Homo sapi
c	2	48	12.1	143372	9	AL137847	AL137847 Human DNA
c	3	40	10.1	123779	30	AC021025	Ac021025 Homo sapi
	4	40	10.1	128118	2	AC076969	AC076969 Homo sapi
c	5	40	10.1	128583	9	AC121249	AC121249 Homo sapi
c	6	40	10.1	148290	9	AC107979	AC107979 Homo sapi
c	7	40	10.1	165649	9	AC103996	AC103996 Homo sapi
	8	40	10.1	178650	9	AC104303	AC104303 Homo sapi
c	9	40	10.1	192826	9	AC090762	AC090762 Homo sapi
	10	39	9.8	32918	2	AC007445	AC007445 Homo sapi
	11	39	9.8	38936	9	AL358817	AL358817 Human DNA
	12	39	9.8	124271	2	AC025179	AC025179 Homo sapi
	13	39	9.8	146671	9	AC008814	AC008814 Homo sapi
c	14	39	9.8	159747	2	AP001019	AP001019 Homo sapi
	15	39	9.8	162740	2	AC034249	AC034249 Homo sapi
c	16	39	9.8	169772	9	AC069538	AC069538 Homo sapi
c	17	39	9.8	175466	9	AL607077	AL607077 Human DNA
c	18	38	9.6	110000	2	AL691517_2	Continuation (3 of
	19	38	9.6	166706	9	AC068875	AC068875 Homo sapi
	20	38	9.6	207408	2	AC068618	AC068618 Homo sapi
c	21	38	9.6	207548	9	AC087283	AC087283 Homo sapi
c	22	37	9.3	135090	9	HS164L12	AL009028 Homo sapi
c	23	36	9.1	121720	9	AL591491	AL591491 Human DNA
c	24	35	8.8	153940	9	AC022294	AC022294 Homo sapi
c	25	35	8.8	176253	2	AP001078	AP001078 Homo sapi
c	26	35	8.8	205639	9	AP001793	AP001793 Homo sapi
	27	35	8.8	210734	2	AC015676	AC015676 Homo sapi
c	28	35	8.8	212055	2	AP000899	AP000899 Homo sapi
	29	35	8.8	325069	2	AC079737	AC079737 Homo sapi
	30	34	8.6	99577	9	AC026324	AC026324 Homo sapi
	31	34	8.6	108040	2	AC068150	AC068150 Homo sapi
c	32	34	8.6	134760	9	AC099484	AC099484 Homo sapi
c	33	34	8.6	146059	2	AC019030	AC019030 Homo sapi
c	34	34	8.6	171347	9	AC099776	AC099776 Homo sapi
	35	34	8.6	172206	9	AC092119	AC092119 Homo sapi
c	36	34	8.6	172567	2	AC015493	AC015493 Homo sapi
	37	34	8.6	173166	9	AC092375	AC092375 Homo sapi
	38	34	8.6	177447	2	AC104687	AC104687 Homo sapi
c	39	34	8.6	182881	9	AC090980	AC090980 Homo sapi
	40	34	8.6	183016	2	AC068611	AC068611 Homo sapi
c	41	34	8.6	193929	9	AC090797	AC090797 Homo sapi
c	42	34	8.6	273807	2	AC025421	AC025421 Homo sapi
	43	34	8.6	316296	2	AC092285	AC092285 Homo sapi
c	44	33	8.3	33458	9	HSDJ6010I	AL109656 Human DNA
c	45	33	8.3	49616	9	AL365267	AL365267 Human DNA

ALIGNMENTS

RESULT 1
AC012674/c
LOCUS AC012674 169620 bp DNA linear HTG 07-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP1-458H3, WORKING DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AC012674
VERSION AC012674.10 GI:9719580
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169620)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,

Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169620)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Aug 7, 2000 this sequence version replaced gi:8705345.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HMQG
Center clone name: RP1-458H3
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 139025 bases at least Q40
Consensus quality: 154842 bases at least Q30
Consensus quality: 159725 bases at least Q20
Estimated insert size: 162720; sum-of-contigs estimation
Estimated insert size: 171608; agarose-fp estimation
Quality coverage: 3.9x in Q20 bases; agarose-fp estimation
Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 28689: contig of 28689 bp in length
* 28690 28789: gap of unknown length
* 28790 50832: contig of 22043 bp in length
* 50833 50932: gap of unknown length
* 50933 69144: contig of 18212 bp in length
* 69145 69244: gap of unknown length
* 69245 84204: contig of 14960 bp in length
* 84205 84304: gap of unknown length
* 84305 94667: contig of 10363 bp in length
* 94668 94767: gap of unknown length
* 94768 107261: contig of 12494 bp in length
* 107262 107361: gap of unknown length
* 107362 117550: contig of 10189 bp in length
* 117551 117650: gap of unknown length
* 117651 126939: contig of 9289 bp in length
* 126940 127039: gap of unknown length
* 127040 135040: contig of 8001 bp in length
* 135041 135140: gap of unknown length
* 135141 141639: contig of 6499 bp in length
* 141640 141739: gap of unknown length
* 141740 149558: contig of 7819 bp in length
* 149559 149658: gap of unknown length
* 149659 154562: contig of 4904 bp in length
* 154563 154662: gap of unknown length
* 154663 158987: contig of 4325 bp in length
* 158988 159087: gap of unknown length
* 159088 162376: contig of 3289 bp in length
* 162377 162476: gap of unknown length
* 162477 165191: contig of 2715 bp in length
* 165192 165291: gap of unknown length
* 165292 167173: contig of 1882 bp in length
* 167174 167273: gap of unknown length
* 167274 168393: contig of 1120 bp in length
* 168394 168493: gap of unknown length
* 168494 169620: contig of 1127 bp in length.

FEATURES Location/Qualifiers
source 1. .169620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP1-458H3"

BASE COUNT 52024 a 33180 c 32128 g 50322 t 1966 others
ORIGIN

Query Match 75.5%; Score 299; DB 2; Length 169620;
Best Local Similarity 99.7%; Pred. No. 3.2e-170;
Matches 349; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87441 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 87382

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87381 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 87322

Qy    121 CATACACAGCTCTGTATATCGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 180
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  87321 CATACACAGCTCTGTATATCGGACCCAAGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 87262

Qy    181 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87261 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 87202

Qy    241 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87201 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 87142

Qy    301 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAG 350
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87141 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAG 87092
```

RESULT 2
AL137847/c
LOCUS AL137847 143372 bp DNA linear PRI 16-NOV-2001
DEFINITION Human DNA sequence from clone RP11-439K3 on chromosome 9q22.2-31.1, complete sequence.
ACCESSION AL137847
VERSION AL137847.12 GI:16973786
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143372)
AUTHORS Kimberley,A.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 17, 2001 this sequence version replaced gi:16408610.

```
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em:, EMBL; Sw:,
          SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr9
          RP11-439K3 is from the library RPCI-11.2 constructed by the group
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-439K3 It may be shorter because we sequence overlapping
          sections only once, except for a short overlap.
          The true left end of clone RP11-439K3 is at 1 in this sequence. The
          true left end of clone RP11-344L17 is at 141373 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .143372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-439K3"
                     /clone_lib="RPCI-11.2"
BASE COUNT    42809 a  30224 c  28152 g  42187 t
ORIGIN

  Query Match             12.1%;  Score 48;  DB 9;  Length 143372;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-17;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||||||||||
    82306 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 82259


RESULT 3
AC021025/c
ID   AC021025   standard; DNA; HTG; 123779 BP.
XX
AC   AC021025;
XX
SV   AC021025.9
XX
DT   14-JAN-2000 (Rel. 62, Created)
DT   08-AUG-2000 (Rel. 64, Last updated, Version 11)
XX
DE   Homo sapiens chromosome 3 clone RP11-79K17, WORKING DRAFT SEQUENCE, 10
DE   unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-123779
RA   Muzny D.M., Adams C., Bailey M., Barbaria J., Blankenburg K., Bodota B.,
RA   Bouck J., Bowie S., Brooks A., Buhay C., Bunac C., Burkett C., Burrows J.,
RA   Carter M., Chacko J., Chen Z., Cox C., David R., Delgado O., Deshazo D.,
RA   Ding Y., Domah-Rashid N., Dugan-Rocha S., Durbin K.J., Fernandez C.,
RA   Ferraguto D., Forcum-Tansey J., Frantz P., Ganesh R., Gorrell J.H.,
RA   Gorrell L.L., Guevara W., Harris K., Hernandez J., Hodgson A., Hogues M.,
RA   Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M.,
RA   Kelly S., Kondejewski N., Kong Y., Kovar C., Leal B., Li Z., Lichtarge O.,
RA   Liu J., Liu W., Logan O., Lozado R.J., Lu J., Lucier R., Martin R.,
RA   Martinez C., McLeod M.P., Mei G., Morgan M., Morris S., Nash S., Nelson A.,
RA   Nguyen R., Nguyen N., Nguyen S., Oswal G., Parish B., Paxton S., Payton B.,
RA   Perez L., Pu L.L., Quiles M., Reiter D., Rives M., Samuel S., Say J.,
RA   Scherer S., Shah E., Shen H., Simon M., Sparks A., Stamps A., Sucgang R.,
RA   Tabor P., Taylor T., Vasquez L., Vinson R., Vo Q., Wahbah M.,
RA   Watlington S., Weinstock G., Weinstock I.R., Williamson A., Worley K.,
RA   Wren J., Wrensford G., Yu W., Zhou X., Nelson D., Gibbs R.;
RT   "Direct Submission";
RL   Unpublished.
XX
RN   [2]
RP   1-123779
RA   Worley K.C.;
RT   ;
RL   Submitted (12-JAN-2000) to the EMBL/GenBank/DDBJ databases.
RL   Human Genome Sequencing Center, Department of Molecular and Human Genetics,
RL   Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
XX
CC   On Aug 7, 2000 this sequence version replaced gi:8700003.
CC   -------------- Genome Center
CC   Center: Baylor College of Medicine
CC   Center code: BCM
CC   Web site: http://www.hgsc.bcm.tmc.edu/
CC   Contact: hgsc-help@bcm.tmc.edu
CC   -------------- Project Information
CC   Center project name: HMXB
CC   Center clone name: RP11-79K17
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; L08821
CC   Chemistry: Dye-terminator Big Dye: 100% of reads
CC   Assembly program: Phrap; version 0.990329
CC   Consensus quality: 112332 bases at least Q40
CC   Consensus quality: 117272 bases at least Q30
CC   Consensus quality: 119789 bases at least Q20
CC   Estimated insert size: 121963; sum-of-contigs estimation
CC   Quality coverage: 0x in Q20 bases; agarose-fp estimation
CC   Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation
CC   --------------
CC   * NOTE: Estimated insert size may differ from sequence length
CC   *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 10 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1    35931: contig of 35931 bp in length
CC   *    35932    36031: gap of unknown length
CC   *    36032    60754: contig of 24723 bp in length
CC   *    60755    60854: gap of unknown length
CC   *    60855    71667: contig of 10813 bp in length
CC   *    71668    71767: gap of unknown length
CC   *    71768    84826: contig of 13059 bp in length
CC   *    84827    84926: gap of unknown length
CC   *    84927    96721: contig of 11795 bp in length
CC   *    96722    96821: gap of unknown length
CC   *    96822   105594: contig of 8773 bp in length
CC   *   105595   105694: gap of unknown length
CC   *   105695   112907: contig of 7213 bp in length
CC   *   112908   113007: gap of unknown length
CC   *   113008   120198: contig of 7191 bp in length
CC   *   120199   120298: gap of unknown length
```

```
CC   *   120299   122220: contig of 1922 bp in length
CC   *   122221   122320: gap of unknown length
CC   *   122321   123779: contig of 1459 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .123779
FT                   /chromosome="3"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-79K17"
XX
SQ   Sequence 123779 BP; 37802 A; 24017 C; 23437 G; 37612 T; 911 other;

  Query Match          10.1%;  Score 40;  DB 30;  Length 123779;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
    3083 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 3044


RESULT 4
AC076969
LOCUS       AC076969              128118 bp    DNA     linear   HTG 15-OCT-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K12, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC076969
VERSION     AC076969.6  GI:16117967
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128118)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128118)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:10047573.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBRJ
                Center clone name: RP11-79K12
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 111307 bases at least Q40
                Consensus quality: 117905 bases at least Q30
                Consensus quality: 121494 bases at least Q20
                Estimated insert size: 122854; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20354: contig of 20354 bp in length
            *    20355    20454: gap of unknown length
            *    20455    33071: contig of 12617 bp in length
            *    33072    33171: gap of unknown length
            *    33172    46935: contig of 13764 bp in length
            *    46936    47035: gap of unknown length
            *    47036    57066: contig of 10031 bp in length
            *    57067    57166: gap of unknown length
            *    57167    70378: contig of 13212 bp in length
            *    70379    70478: gap of unknown length
            *    70479    81786: contig of 11308 bp in length
            *    81787    81886: gap of unknown length
            *    81887    91428: contig of 9542 bp in length
            *    91429    91528: gap of unknown length
            *    91529   100455: contig of 8927 bp in length
            *   100456   100555: gap of unknown length
            *   100556   107879: contig of 7324 bp in length
            *   107880   107979: gap of unknown length
            *   107980   114149: contig of 6170 bp in length
            *   114150   114249: gap of unknown length
            *   114250   120186: contig of 5937 bp in length
            *   120187   120286: gap of unknown length
            *   120287   123236: contig of 2950 bp in length
            *   123237   123336: gap of unknown length
            *   123337   125606: contig of 2270 bp in length
            *   125607   125706: gap of unknown length
            *   125707   128118: contig of 2412 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128118
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K12"
```

```
BASE COUNT    35433 a  25053 c  23792 g  42367 t   1473 others
ORIGIN

  Query Match           10.1%;  Score 40;  DB 2;  Length 128118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db   55106 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 55145


RESULT 5
AC121249/c
LOCUS       AC121249              128583 bp    DNA     linear   PRI 01-JUN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K17, complete sequence.
ACCESSION   AC121249 AC021025
VERSION     AC121249.1  GI:20806313
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 16, 2002 this sequence version replaced gi:9719675.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-79K17 (bc0196)
            ------------- Summary Statistics
                Sequencing vector: plasmid; 45% of reads
                Sequencing vector: M13; L08821; 55% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 128548 bases at least Q40
                Consensus quality: 128580 bases at least Q30
                Consensus quality: 128581 bases at least Q20
                Insert size: 128583; sum-of-contigs
                Quality coverage: 13.4x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  RP11-147N17 (UWGC:bc0267) AC104300, 22116-bp overlap
                3':  RP11-391P4 (UWGC:bc0402) AC104303, 114502-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
12049	12136	4077	4013	8696	8902
579	<800	2067	2145	6	<800
1056	1065	9698	9932	1323	1337
13085	13126	3729	3734	1024	961
10944	10836	1632	1616	8421	8408
581	<800	8069	8139	1114	1126
882	887	1507	1485	3538	3534
8435	8467	11939	11871	6051	6022
3924	4103	133	<800	5129	5048
9842	9759	2135	2145	58	<800
4401	4377	78	<800	1895	1903
4458	4377	5873	5861	799	<800
4587	4522	896	902	858	885
6048	5947	3312	3356	931	885
1395	1385	1505	1485	859	885
3059	3071	6694	6716	3934	3936
3911	3861	10032	9932	1065	1058
138	<800	2168	2145	2629	2723
50	<800	1794	1785	859	885
582	<800	227	<800	6855	6888
2225	2229	4002	4013	7821	7819
984	967	2714	2726	7210	7201
887	887	5628	5624	6992	6888

1042	1065	743	758	1204	1224
2783	2808	3045	2963	2722	2723
7074	7158	1288	1262	321	<800
4087	3861	967	973	1066	1058
7976	7926	461	<800	470	<800
1233	1199	4649	4618	3055	3072
3875	3861	5289	5417	10961	10848
27	<800	1098	1095	6725	6888
1137	1065	798	809	1896	1903
750	749	111	<800	8682	8666
214	<800	7644	7609	4438	4342
836	887	10084	9932	514	<800
1670	1651	3641	3589	3077	3072
2650	2609	7555	7609	1371	1337
1890	1872			4185	4087
4615	4522			719	<800
1321	1273			4130	4087
				367	<800
				2979	3072
				333	<800

```
FEATURES             Location/Qualifiers
     source          1. .128583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K17"
                     /clone_lib="RPCI human BAC library 11"
     COUNT     37777 a  24752 c  24789 g  41265 t
    IN

  Query Match             10.1%;  Score 40;  DB 9;  Length 128583;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db   27283 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 27244


RESULT 6
AC107979/c
LOCUS       AC107979              148290 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone CTD-3049M7, complete sequence.
ACCESSION   AC107979
VERSION     AC107979.7  GI:21592043
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-3049M7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 26, 2002 this sequence version replaced gi:21321840.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24533
                Center clone name: 3049_M_7
            --------------.
FEATURES             Location/Qualifiers
     source          1. .148290
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-3049M7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   complement(1982. .2092)
                     /rpt_family="MIR"
     repeat_region   complement(3356. .3654)
                     /rpt_family="AluSx"
     repeat_region   3789. .3844
                     /rpt_family="MER58"
     repeat_region   complement(4429. .4728)
                     /rpt_family="AluSp"
     repeat_region   5494. .5723
                     /rpt_family="(TA)n"
     repeat_region   5727. .5766
                     /rpt_family="(GA)n"
     repeat_region   5806. .5857
                     /rpt_family="AT_rich"
     repeat_region   6377. .6581
                     /rpt_family="L1"
     repeat_region   6647. .7069
                     /rpt_family="L1ME3"
     repeat_region   7057. .7631
                     /rpt_family="L1ME3"
     repeat_region   complement(8821. .8958)
                     /rpt_family="MIR3"
     repeat_region   11129. .11224
                     /rpt_family="FRAM"
     repeat_region   11229. .11338
                     /rpt_family="CT-rich"
     repeat_region   11943. .12090
                     /rpt_family="MIR"
     repeat_region   12818. .13060
                     /rpt_family="MIR"
     repeat_region   13118. .19146
                     /rpt_family="L1PA2"
     repeat_region   20846. .20912
                     /rpt_family="CT-rich"
     repeat_region   21294. .21314
                     /rpt_family="AT_rich"
     repeat_region   21635. .23417
                     /rpt_family="L1PA16"
     repeat_region   23438. .25940
                     /rpt_family="L1PA16"
     repeat_region   26121. .26145
                     /rpt_family="(TTCA)n"
     repeat_region   26293. .26417
                     /rpt_family="MIR"
     repeat_region   complement(26872. .26937)
                     /rpt_family="L2"
     repeat_region   complement(27064. .27431)
                     /rpt_family="L2"
     repeat_region   27917. .28357
                     /rpt_family="MLT1B"
     repeat_region   28460. .28806
                     /rpt_family="MLT1I"
     repeat_region   29509. .29698
                     /rpt_family="MIR"
     repeat_region   29716. .29790
                     /rpt_family="CT-rich"
     repeat_region   29791. .29811
                     /rpt_family="AT_rich"
     repeat_region   complement(31182. .31787)
                     /rpt_family="L2"
     repeat_region   complement(32398. .32581)
                     /rpt_family="AluSq"
     repeat_region   complement(32582. .32777)
                     /rpt_family="L1PB1"
     repeat_region   complement(32778. .32912)
                     /rpt_family="AluSq"
     repeat_region   33104. .33207
                     /rpt_family="L1MC5"
     repeat_region   complement(33208. .33365)
                     /rpt_family="MER53"
     repeat_region   33366. .33529
                     /rpt_family="L1MC5"
     repeat_region   33877. .34506
                     /rpt_family="L1MEd"
     repeat_region   34667. .34808
                     /rpt_family="L1MEd"
     repeat_region   35585. .36698
                     /rpt_family="L1M4"
     repeat_region   complement(36699. .36830)
                     /rpt_family="FLAM_C"
     repeat_region   36831. .38141
                     /rpt_family="L1M4"
     repeat_region   38194. .38373
                     /rpt_family="MLT1I"
     repeat_region   39530. .39757
                     /rpt_family="MIR"
     repeat_region   39992. .40017
                     /rpt_family="(TTTTG)n"
     repeat_region   40283. .40583
                     /rpt_family="AluY"
     repeat_region   41414. .41616
                     /rpt_family="MLT1C"

  Query Match           10.1%;  Score 40;  DB 9;  Length 148290;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           |||||||||||||||||||||||||||||||||||||||||
Db   84499 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 84460


RESULT 7
AC103996/c
LOCUS       AC103996              165649 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-76E17, complete sequence.
ACCESSION   AC103996
VERSION     AC103996.7  GI:21637504
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-76E17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165649)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
```

Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 165649)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 165649)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jul 1, 2002 this sequence version replaced gi:21592191.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L21917
Center clone name: 76_E_17
---------- .

FEATURES	Location/Qualifiers
source	1. .165649
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="15"
	/map="15"
	/clone="RP11-76E17"
	/clone_lib="RPCI-11 Human Male BAC"
repeat_region	complement(2. .865)
	/rpt_family="L1PA13"
repeat_region	complement(864. .894)
	/rpt_family="L1PA13"
repeat_region	complement(895. .1248)
	/rpt_family="THE1A"
repeat_region	complement(1249. .1504)
	/rpt_family="L1PA13"
repeat_region	1505. .1626
	/rpt_family="AluSx"
repeat_region	1627. .1659
	/rpt_family="(CAAA)n"
repeat_region	1660. .1831
	/rpt_family="AluSx"
repeat_region	complement(1832. .3132)
	/rpt_family="L1PA13"
repeat_region	complement(3140. .3699)
	/rpt_family="L1M4c"
repeat_region	complement(4461. .4755)
	/rpt_family="AluSx"
repeat_region	5027. .5397
	/rpt_family="L2"
repeat_region	complement(5650. .6330)
	/rpt_family="L1ME1"
repeat_region	complement(6343. .6460)
	/rpt_family="L1ME1"
repeat_region	7214. .7319
	/rpt_family="L1MC3"
repeat_region	7354. .7655
	/rpt_family="L1MC3"
repeat_region	7659. .7743
	/rpt_family="L1PA10"
repeat_region	complement(7745. .7843)
	/rpt_family="AluSp/q"
repeat_region	7844. .14156
	/rpt_family="L1PA10"
repeat_region	14198. .14328
	/rpt_family="AluY"
repeat_region	14340. .14503
	/rpt_family="(TA)n"
repeat_region	14507. .15032
	/rpt_family="L1MC3"
repeat_region	15036. .15220
	/rpt_family="L1MC3"
repeat_region	complement(15218. .16449)
	/rpt_family="L1PA4"
repeat_region	16450. .18154
	/rpt_family="L1PA4"
repeat_region	18155. .19210
	/rpt_family="L1MC3"
repeat_region	19211. .19266
	/rpt_family="(TA)n"
repeat_region	19267. .19372
	/rpt_family="L1MC3"
repeat_region	19401. .19476
	/rpt_family="(TTATA)n"
repeat_region	19483. .19546
	/rpt_family="(CATATA)n"
repeat_region	19596. .19625
	/rpt_family="AT_rich"

```
     repeat_region   19696. .19750
                     /rpt_family="GA-rich"
     repeat_region   19752. .19920
                     /rpt_family="L1MD3"
     repeat_region   19994. .20102
                     /rpt_family="L2"
     unsure          20049. .20114
                     /note="single clone coverage"
     repeat_region   20485. .20655
                     /rpt_family="MIR3"
     repeat_region   complement(21285. .21441)
                     /rpt_family="MIR"
     repeat_region   21496. .21717
                     /rpt_family="L2"
     repeat_region   complement(21720. .21796)
                     /rpt_family="MIR"
     repeat_region   complement(21943. .22131)
                     /rpt_family="MIR"
     repeat_region   23082. .23195
                     /rpt_family="L2"
     repeat_region   23198. .23248
                     /rpt_family="GA-rich"
     repeat_region   23267. .23333
                     /rpt_family="(CATA)n"
     repeat_region   complement(24465. .24833)
                     /rpt_family="MLT1A2"
     repeat_region   26142. .26334
                     /rpt_family="MIR"
     repeat_region   complement(27659. .27811)
                     /rpt_family="MIR"
     repeat_region   complement(28811. .28880)
                     /rpt_family="MIR"
     repeat_region   complement(29026. .29269)
                     /rpt_family="MIR"
     repeat_region   complement(29673. .29706)
                     /rpt_family="MSTB"

  Query Match          10.1%;  Score 40;  DB 9;  Length 165649;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           ||||||||||||||||||||||||||||||||||||||||
Db    8110 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 8071


RESULT 8
AC104303
LOCUS       AC104303              178650 bp    DNA     linear   PRI 25-FEB-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-391P4, complete sequence.
ACCESSION   AC104303 AC064830
VERSION     AC104303.2  GI:18874945
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 178650)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Feb 25, 2002 this sequence version replaced gi:17402782.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: WUGSC
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-391P4 (bc0402)
            ------------- Summary Statistics
                Sequencing vector: unknown; 55% of reads
                Sequencing vector: plasmid; 45% of reads
                Chemistry: Dye-terminator ET; 89% of reads
                Chemistry: Dye-terminator Big Dye; 11% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 178494 bases at least Q40
                Consensus quality: 178631 bases at least Q30
                Consensus quality: 178650 bases at least Q20
                Insert size: 178648; sum-of-contigs
                Quality coverage: 8.0x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  RP11-475O23 (UWGC:bc0439) AC023346
                3':  RP11-79K12 AC076969
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                   BglII                EcoRI               HindIII
            ------------------   ------------------   ------------------
            SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
            ---------  --------  ---------  --------  ---------  --------

            ---------            ---------            ---------
                 7050       7417       8696       8715       7163       7098
            ---------            ---------            ---------
                 2067       2138          6       <800       6382       6518
            ---------            ---------            ---------
                13472      13264       3077       3083        512       <800
            ---------            ---------            ---------
                 7644       7859        514       <800        449       <800
            ---------            ---------            ---------
                  111       <800       4438       4377       7988       7884
            ---------            ---------            ---------
                  798        783       8682       8715      11779      11698
            ---------            ---------            ---------
```

1098	1109	1896	1876	472	<800
5288	5395	6724	6948	783	787
4649	4507	10961	10773	926	926
461	<800	3055	3083	2431	2538
967	982	470	<800	1015	1029
1288	1268	1066	1037	758	<800
3045	2923	321	<800	1550	1541
743	783	2722	2683	361	<800
5628	5682	1204	1300	763	<800
2714	2699	6992	7281	2975	3002
4002	3900	7210	7830	3299	3468
227	<800	7821	8101	1907	1894
1794	1784	6855	6948	4020	4220
2168	2138	859	857	5158	5126
10032	9744	2629	2683	497	<800
6694	6810	1065	1037	4321	4220
1505	1491	3934	3965	887	926
3312	3304	859	857	180	<800
896	912	931	935	4969	4887
5873	5938	858	857	4755	4674
78	<800	799	857	26	<800
2135	2138	1895	1876	906	926
133	<800	58	<800	2325	2346
11939	11575	5129	5076	369	<800
1507	1491	6051	6048	2523	2538
8069	8181	3538	3537	1889	1894
1632	1617	1114	1094	4437	4428
3729	3678	8420	8391	373	<800
5553	5395	1024	1037	24	<800
358	<800	1323	1380	263	<800
3764	3678	1502	1481	632	<800
7022	6810	4538	4539	1258	1247
8534	8821	1414	1380	1114	1083
5206	5050	521	<800	1370	1320
9052	9744	8346	8391	1262	1247
1625	1617	39	<800	1546	1541
749	783	2716	2683	459	<800
10190	9744	2546	2683	135	<800
3857	3795	1064	1037	3142	3183
6144	6490	926	935	109	<800
514	<800	5154	5076	7858	7884
1364	1342	1975	1981	1891	1894
669	<800	628	<800	723	<800
		2398	2399	6975	7098
		8148	8391	1878	1894
		1713	1673	3747	4012
		135	<800	3146	3183
		566	<800	3450	3737
		1115	1094	390	<800
		2344	2399	560	<800
		1119	1188	7249	7098

Query Match 10.1%; Score 40; DB 9; Length 178650;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||||||||||||||||||||||||
Db  165449 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 165488
```

```
RESULT 9
AC090762/c
LOCUS       AC090762              192826 bp    DNA     linear   PRI 28-FEB-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-387E8, complete sequence.
ACCESSION   AC090762
VERSION     AC090762.9  GI:18997378
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-387E8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
```

```
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2002 this sequence version replaced gi:18377189.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12392
                Center clone name: 387_E_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192826
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-387E8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(864. .1222)
                     /rpt_family="L1PB3"
     repeat_region   1223. .1252
                     /rpt_family="(TAGG)n"
     repeat_region   complement(1253. .1492)
                     /rpt_family="L1PB3"
     repeat_region   4287. .4576
                     /rpt_family="AluJb"
     repeat_region   4671. .5034
                     /rpt_family="THE1B"
     repeat_region   5130. .5438
                     /rpt_family="AluSx"
     repeat_region   5517. .5662
                     /rpt_family="MIR"
     repeat_region   5798. .5827
                     /rpt_family="AT_rich"
     repeat_region   5985. .6114
                     /rpt_family="AluJb"
     repeat_region   6227. .6253
                     /rpt_family="(CA)n"
     repeat_region   7329. .7446
                     /rpt_family="MIR"
     repeat_region   complement(8147. .8452)
                     /rpt_family="AluJo"
     repeat_region   8540. .8639
                     /rpt_family="MER45"
     repeat_region   complement(8832. .8916)
                     /rpt_family="L2"
     repeat_region   10391. .10599
                     /rpt_family="MER3"
     repeat_region   11688. .11828
                     /rpt_family="L1MC/D"
     repeat_region   12029. .12078
                     /rpt_family="AT_rich"
     repeat_region   12092. .12447
                     /rpt_family="THE1C"
     repeat_region   complement(13616. .13751)
                     /rpt_family="MIR3"
     repeat_region   complement(13958. .14142)
                     /rpt_family="MIR"
     repeat_region   complement(15277. .15553)
                     /rpt_family="MER8"
     repeat_region   complement(15811. .16005)
                     /rpt_family="MIR"
     repeat_region   complement(16101. .16440)
                     /rpt_family="L3"
     repeat_region   16920. .16958
                     /rpt_family="(TCCC)n"
     repeat_region   complement(17145. .17444)
                     /rpt_family="AluSx"
     repeat_region   complement(18418. .19953)
                     /rpt_family="L1MEc"
     repeat_region   complement(19978. .20262)
                     /rpt_family="L1MEc"
     repeat_region   complement(20288. .20794)
                     /rpt_family="L1MEc"
     repeat_region   complement(20822. .21097)
                     /rpt_family="L1MEc"
     repeat_region   complement(21345. .21743)
                     /rpt_family="L1MEc"
     repeat_region   22599. .23518
                     /rpt_family="L1MEc"
     repeat_region   23527. .23901
                     /rpt_family="L1MEc"
     repeat_region   complement(23927. .24026)
                     /rpt_family="MSTB1"
     repeat_region   complement(24027. .24256)
                     /rpt_family="MER30"
     repeat_region   complement(24257. .24557)
                     /rpt_family="MSTB1"
     repeat_region   24563. .24594
                     /rpt_family="AT_rich"
     repeat_region   complement(25258. .25639)
                     /rpt_family="L1MCc"
     repeat_region   complement(25688. .25838)
                     /rpt_family="L1MCc"
     repeat_region   26167. .26506
                     /rpt_family="THE1B"
     repeat_region   26746. .27094
                     /rpt_family="Tigger2a"
     repeat_region   27095. .27184
                     /rpt_family="MADE1"
     repeat_region   27185. .27279
                     /rpt_family="Tigger2a"
     repeat_region   28321. .28363
                     /rpt_family="(TG)n"
     repeat_region   29172. .29333
                     /rpt_family="MIR"
     repeat_region   30664. .30898
                     /rpt_family="L1MB8"
     repeat_region   31577. .31598
                     /rpt_family="AT_rich"
     repeat_region   32378. .32475
```

```
                     /rpt_family="CT-rich"
     repeat_region   34655. .34960
                     /rpt_family="AluY"
     repeat_region   complement(35296. .35389)
                     /rpt_family="MER5B"
     repeat_region   complement(35419. .35517)
                     /rpt_family="L1MC4"
     repeat_region   complement(35518. .35823)
                     /rpt_family="AluSx"
     repeat_region   complement(35824. .36030)
                     /rpt_family="L1MC4"
     repeat_region   36682. .36758
                     /rpt_family="MER5A"
     repeat_region   36975. .37160
                     /rpt_family="MER5A"
     repeat_region   38140. .38353
                     /rpt_family="MER96B"
     repeat_region   38633. .38656
                     /rpt_family="(TTTA)n"
     repeat_region   complement(38657. .40056)
                     /rpt_family="L1PA5"
     repeat_region   complement(40060. .40250)

  Query Match             10.1%;  Score 40;  DB 9;  Length 192826;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-12;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 107
           |||||||||||||||||||||||||||||||||||||||||
Db 190946 TAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTC 190907


RESULT 10
AC007445
LOCUS       AC007445               32918 bp    DNA     linear   HTG 30-JUN-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-344B7 map 18, *** SEQUENCING
            IN PROGRESS ***, 1 ordered piece.
ACCESSION   AC007445
VERSION     AC007445.5  GI:8844149
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-344B7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32918)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 30, 2000 this sequence version replaced gi:8705092.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L571
                Center clone name: 344_B_7
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    32918: contig of 32918 bp in length.
FEATURES             Location/Qualifiers
     source          1. .32918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-344B7"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     9462 a   6493 c   6865 g   9853 t    245 others
ORIGIN

  Query Match              9.8%;  Score 39;  DB 2;  Length 32918;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db    3535 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 3573


RESULT 11
AL358817
LOCUS       AL358817               38936 bp    DNA     linear   PRI 06-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-399N22 on chromosome 10,
            complete sequence.
ACCESSION   AL358817
VERSION     AL358817.18  GI:15990637
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38936)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2001 this sequence version replaced gi:14669268.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
```

```
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-399N22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-399N22 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-432J9 is at 36937 in this sequence.
            The true right end of clone RP11-91A1 is at 2000 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .38936
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-399N22"
                     /clone_lib="RPCI-11.2"
BASE COUNT     9315 a   9079 c   9111 g  11431 t
ORIGIN

  Query Match             9.8%;  Score 39;  DB 9;  Length 38936;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db   28896 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 28934


RESULT 12
AC025179
LOCUS        AC025179              124271 bp    DNA     linear   HTG 20-APR-2001
DEFINITION   Homo sapiens chromosome 5 clone CTD-2174B5, WORKING DRAFT SEQUENCE,
             8 unordered pieces.
ACCESSION    AC025179
VERSION      AC025179.4  GI:13699647
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE       Homo sapiens.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 124271)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Sequencing of Human Chromosome 5
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 124271)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      On Apr 20, 2001 this sequence version replaced gi:7711794.
             -------Genome Center
             Center: Joint Genome Institute
             Center Code: JGI
             Web site: http://www.jgi.doe.gov
             -------
             Project Information
             Center Project Name: 694394
             Center clone name: CITB-H1_2174B5
             -------
             Summary Statistics
             Consensus quality: 116609 bases at least Q40
             Consensus quality: 120226 bases at least Q30
             Consensus quality: 121274 bases at least Q20
             Estimated insert size: 117160; agarose-fp estimation
             Estimated insert size: 123571; sum-of-contigs estimation
             Quality coverage:  6.46 in Q20 bases; agarose-fp estimation
             Quality coverage:  6.12 in Q20 bases; sum-of-contigs estimation.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 8 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     1052: contig of 1052 bp in length
             *    1053     1152: gap of unknown length
             *    1153     3331: contig of 2179 bp in length
             *    3332     3431: gap of unknown length
             *    3432     7370: contig of 3939 bp in length
             *    7371     7470: gap of unknown length
             *    7471    15483: contig of 8013 bp in length
             *   15484    15583: gap of unknown length
             *   15584    24916: contig of 9333 bp in length
             *   24917    25016: gap of unknown length
             *   25017    39922: contig of 14906 bp in length
             *   39923    40022: gap of unknown length
             *   40023    68684: contig of 28662 bp in length
             *   68685    68784: gap of unknown length
             *   68785   124271: contig of 55487 bp in length.
FEATURES             Location/Qualifiers
     source          1. .124271
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2174B5"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT     36863 a  24233 c  23599 g  38876 t    700 others
ORIGIN

  Query Match             9.8%;  Score 39;  DB 2;  Length 124271;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
           |||||||||||||||||||||||||||||||||||||||
Db   79646 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 79684


RESULT 13
AC008814
LOCUS        AC008814              146671 bp    DNA     linear   PRI 31-OCT-2001
DEFINITION   Homo sapiens chromosome 5 clone CTD-2117L12, complete sequence.
ACCESSION    AC008814
VERSION      AC008814.6  GI:16554342
KEYWORDS     HTG.
SOURCE       Homo sapiens.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 146671)
  AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 146671)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3  (bases 1 to 146671)
  AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE      Direct Submission
  JOURNAL    Submitted (31-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
COMMENT      On Oct 31, 2001 this sequence version replaced gi:15290309.
             Draft Sequence Produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing Completed at Stanford Human Genome Center
             www-shgc.stanford.edu
             Quality: Phrap Quality >=40 99.3% of Sequence;
             Estimated Total Number of Errors is 0.7.
```

```
FEATURES             Location/Qualifiers
     source          1. .146671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2117L12"
BASE COUNT    44951 a  28815 c  28434 g  44471 t
ORIGIN

  Query Match             9.8%;  Score 39;  DB 9;  Length 146671;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
           |||||||||||||||||||||||||||||||||||||||
Db   54480 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 54518


RESULT 14
AP001019/c
LOCUS       AP001019              159747 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-752I11 map 18p11.3, WORKING
            DRAFT SEQUENCE, 24 unordered pieces.
ACCESSION   AP001019
VERSION     AP001019.2  GI:8117689
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-752I11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159747)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 159,747 genomic DNA of 18p11.3
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 159747)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997769.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-752I11
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 145356 bases at least Q40
                Consensus quality: 152227 bases at least Q30
                Consensus quality: 155843 bases at least Q20
                Insert size: 157447; sum-of-contigs
                Quality coverage: 4.51x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            24 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     18444 contig of  18444 bp in length
                18545     29631 contig of  11087 bp in length
                29732     43143 contig of  13412 bp in length
                43244     57170 contig of  13927 bp in length
                57271     68240 contig of  10970 bp in length
                68341     78134 contig of   9794 bp in length
                78235     87873 contig of   9639 bp in length
                87974     96197 contig of   8224 bp in length
                96298    103976 contig of   7679 bp in length
               104077    110726 contig of   6650 bp in length
               110827    116866 contig of   6040 bp in length
               116967    122789 contig of   5823 bp in length
               122890    127665 contig of   4776 bp in length
               127766    132618 contig of   4853 bp in length
               132719    136051 contig of   3333 bp in length
               136152    139749 contig of   3598 bp in length
               139850    142810 contig of   2961 bp in length
               142911    145855 contig of   2945 bp in length
               145956    148840 contig of   2885 bp in length
               148941    151243 contig of   2303 bp in length
               151344    153454 contig of   2111 bp in length
               153555    156134 contig of   2580 bp in length
               156235    158089 contig of   1855 bp in length
               158190    159747 contig of   1558 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    18444: contig of 18444 bp in length
*    18445 18544: gap of      100 bp
*    18545    29631: contig of 11087 bp in length
*    29632 29731: gap of      100 bp
*    29732    43143: contig of 13412 bp in length
*    43144 43243: gap of      100 bp
*    43244    57170: contig of 13927 bp in length
*    57171 57270: gap of      100 bp
*    57271    68240: contig of 10970 bp in length
*    68241 68340: gap of      100 bp
*    68341    78134: contig of 9794 bp in length
*    78135 78234: gap of      100 bp
*    78235    87873: contig of 9639 bp in length
*    87874 87973: gap of      100 bp
*    87974    96197: contig of 8224 bp in length
*    96198 96297: gap of      100 bp
*    96298   103976: contig of 7679 bp in length
*   103977 104076: gap of      100 bp
*   104077   110726: contig of 6650 bp in length
*   110727 110826: gap of      100 bp
*   110827   116866: contig of 6040 bp in length
*   116867 116966: gap of      100 bp
*   116967   122789: contig of 5823 bp in length
*   122790 122889: gap of      100 bp
*   122890   127665: contig of 4776 bp in length
*   127666 127765: gap of      100 bp
*   127766   132618: contig of 4853 bp in length
*   132619 132718: gap of      100 bp
*   132719   136051: contig of 3333 bp in length
*   136052 136151: gap of      100 bp
*   136152   139749: contig of 3598 bp in length
*   139750 139849: gap of      100 bp
*   139850   142810: contig of 2961 bp in length
*   142811 142910: gap of      100 bp
*   142911   145855: contig of 2945 bp in length
*   145856 145955: gap of      100 bp
*   145956   148840: contig of 2885 bp in length
*   148841 148940: gap of      100 bp
*   148941   151243: contig of 2303 bp in length
*   151244 151343: gap of      100 bp
*   151344   153454: contig of 2111 bp in length
*   153455 153554: gap of      100 bp
```

```
            *   153555   156134: contig of 2580 bp in length
            *   156135 156234: gap of      100 bp
            *   156235   158089: contig of 1855 bp in length
            *   158090 158189: gap of      100 bp
            *   158190   159747: contig of 1558 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159747
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18p11.3"
                     /clone="RP11-752I11"
     misc_feature    1. .18444
                     /note="assembly_fragment"
     misc_feature    18545. .29631
                     /note="assembly_fragment"
     misc_feature    29732. .43143
                     /note="assembly_fragment"
     misc_feature    43244. .57170
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    57271. .68240
                     /note="assembly_fragment"
     misc_feature    68341. .78134
                     /note="assembly_fragment"
     misc_feature    78235. .87873
                     /note="assembly_fragment"
     misc_feature    87974. .96197
                     /note="assembly_fragment"
     misc_feature    96298. .103976
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    104077. .110726
                     /note="assembly_fragment"
     misc_feature    110827. .116866
                     /note="assembly_fragment"
     misc_feature    116967. .122789
                     /note="assembly_fragment"
     misc_feature    122890. .127665
                     /note="assembly_fragment"
     misc_feature    127766. .132618
                     /note="assembly_fragment"
     misc_feature    132719. .136051
                     /note="assembly_fragment"
     misc_feature    136152. .139749
                     /note="assembly_fragment"
     misc_feature    139850. .142810
                     /note="assembly_fragment"
     misc_feature    142911. .145855
                     /note="assembly_fragment"
     misc_feature    145956. .148840
                     /note="assembly_fragment"
     misc_feature    148941. .151243
                     /note="assembly_fragment"
     misc_feature    151344. .153454
                     /note="assembly_fragment"
     misc_feature    153555. .156134
                     /note="assembly_fragment"
     misc_feature    156235. .158089
                     /note="assembly_fragment"
     misc_feature    158190. .159747
                     /note="assembly_fragment"
BASE COUNT    47302 a  30397 c  29897 g  49849 t   2302 others
ORIGIN

  Query Match             9.8%;  Score 39;  DB 2;  Length 159747;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           |||||||||||||||||||||||||||||||||||||||
Db   74932 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 74894


RESULT 15
AC034249
LOCUS       AC034249              162740 bp    DNA     linear   HTG 31-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-427C17, WORKING DRAFT
            SEQUENCE, 6 ordered pieces.
ACCESSION   AC034249
VERSION     AC034249.3  GI:15383785
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162740)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162740)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 31, 2001 this sequence version replaced gi:9211234.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 570398
            Center clone name: RPCI-11_427C17
            -------
            Summary Statistics
            Consensus quality: 158733 bases at least Q40
            Consensus quality: 161469 bases at least Q30
            Consensus quality: 162087 bases at least Q20
            Estimated insert size: 160000; pulse field gel estimation
            Estimated insert size: 162240; sum-of-contigs estimation
            Quality coverage:  9.28 in Q20 bases; pulse field gel estimation
            Quality coverage:  9.16 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    85871: contig of 85871 bp in length
            *    85872    85971: gap of unknown length
            *    85972    96765: contig of 10794 bp in length
            *    96766    96865: gap of unknown length
            *    96866   103302: contig of 6437 bp in length
            *   103303   103402: gap of unknown length
            *   103403   111971: contig of 8569 bp in length
            *   111972   112071: gap of unknown length
            *   112072   129034: contig of 16963 bp in length
            *   129035   129134: gap of unknown length
            *   129135   162740: contig of 33606 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-427C17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    48722 a  31435 c  31563 g  50520 t    500 others
ORIGIN

  Query Match             9.8%;  Score 39;  DB 2;  Length 162740;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-12;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 99
```

```
                    ||||||||||||||||||||||||||||||||||||||||
Db    3599 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGT 3637
```

Search completed: April 25, 2003, 00:41:44
Job time : 1748.12 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:19:38 ; Search time 181.988 Seconds
(without alignments)
4900.271 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_446
Perfect score: 396
Sequence: 1 atgggtggatcttttgcctt.....gamctggatatttcagttga 396

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	396	100.0	447	21	AAC03794	Human secreted pro
	2	396	100.0	447	21	AAZ42680	Human 5' EST isola
	3	29	7.3	966	23	AAS93299	DNA encoding novel
c	4	24	6.1	660	23	AAS73441	DNA encoding novel
	5	22	5.6	3659	7	AAN60204	Interferon-pseudo-
c	6	22	5.6	147724	24	ABK83566	Human cDNA differe
c	7	20	5.1	1179	23	AAS91061	DNA encoding novel
	8	20	5.1	2526	23	AAS66674	DNA encoding novel
c	9	19	4.8	366	22	AAC90702	Human secretory pr
c	10	19	4.8	802	20	AAX84949	Human secreted pro
c	11	19	4.8	1982	21	AAC68089	Human secreted pro
	12	19	4.8	2770	23	ABL09112	Drosophila melanog
c	13	19	4.8	4902	23	ABL28930	Drosophila melanog
	14	18	4.5	31	16	AAT45064	21-hydroxylase fra
c	15	18	4.5	339	21	AAZ43080	C. elegans insulin
	16	18	4.5	431	24	ABK24383	DNA encoding human
	17	18	4.5	431	24	ABK24390	DNA encoding human
c	18	18	4.5	432	22	AAI83964	Human polynucleoti
c	19	18	4.5	570	22	ABA63453	Human foetal liver
c	20	18	4.5	570	22	ABA30652	Probe #9118 for ge
c	21	18	4.5	570	22	AAK11985	Human brain expres
c	22	18	4.5	570	22	AAK37688	Human bone marrow
c	23	18	4.5	570	22	AAI18447	Probe #8380 for ge
c	24	18	4.5	570	22	AAI43563	Probe #12249 used
c	25	18	4.5	570	24	ABS11680	Human genome-deriv
	26	18	4.5	827	23	AAS87115	DNA encoding novel
	27	18	4.5	852	22	AAI94104	Human neuroblastom
	28	18	4.5	870	22	AAI94266	Human neuroblastom
c	29	18	4.5	872	21	AAC79760	Human secreted pro
c	30	18	4.5	3342	23	AAS87118	DNA encoding novel
	31	18	4.5	4886	22	AAK85823	Human immune/haema
	32	18	4.5	6464	22	ABA09665	Human bone marrow
	33	18	4.5	9899	22	AAK85825	Human immune/haema
	34	18	4.5	10918	22	ABA09581	Human bone marrow
	35	18	4.5	10920	24	ABN60009	Novel human coding
c	36	18	4.5	113515	24	ABL34174	Human immune syste
	37	17	4.3	266	20	AAV89709	EST clone CT857.
	38	17	4.3	288	21	AAC70647	Single nucleotide
	39	17	4.3	288	21	AAC70650	Single nucleotide
	40	17	4.3	288	21	AAC70656	Single nucleotide
	41	17	4.3	288	21	AAC70668	Single nucleotide
	42	17	4.3	288	21	AAC70671	Single nucleotide
	43	17	4.3	294	22	AAH73108	Human cervical can
	44	17	4.3	297	22	AAH70408	Human cervical can
	45	17	4.3	318	22	AAH69668	Human cervical can

ALIGNMENTS

RESULT 1
AAC03794
ID AAC03794 standard; cDNA; 447 BP.
XX
AC AAC03794;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3792.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG03788.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3792; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 396; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.8e-183;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 170

Qy    121 CATACACAGCTCTGTATATCGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 CATACACAGCTCTGTATATCGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 230

Qy    181 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 290

Qy    241 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 350

Qy    301 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAGDBGTTCACTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAGDBGTTCACTT 410

Qy    361 GGTCAGTCCATTGTGRGAMCTGGATATTTCAGTTGA 396
          |||||||||||||||||||||||||||||||||||
Db    411 GGTCAGTCCATTGTGRGAMCTGGATATTTCAGTTGA 446
```

RESULT 2
AAZ42680
ID AAZ42680 standard; cDNA; 447 BP.
XX
AC AAZ42680;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:439.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
OS Homo sapiens.
XX
PN WO9953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
DR P-PSDB; AAY65066.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 402; 837pp; English.
XX
CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 396; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.8e-183;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 170

Qy    121 CATACACAGCTCTGTATATCGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 CATACACAGCTCTGTATATCGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCC 230

Qy    181 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CCTGATCTGTGGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACAC 290

Qy    241 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 AATCACTCACTGCTTCCCTTGGCTGGAGGTGGGGCTTTTTTTTTGGCTCCATGCTGCTCC 350

Qy    301 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAGDBGTTCACTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TGGGCGGGCGATTGCCCTACCCCATTTTTCTTCATTCTCTGTGGGTCAAGDBGTTCACTT 410

Qy    361 GGTCAGTCCATTGTGRGAMCTGGATATTTCAGTTGA 396
          |||||||||||||||||||||||||||||||||||
Db    411 GGTCAGTCCATTGTGRGAMCTGGATATTTCAGTTGA 446
```

RESULT 3
AAS93299
ID AAS93299 standard; cDNA; 966 BP.
XX
AC AAS93299;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29103.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG29112.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 29103; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 966 BP; 250 A; 221 C; 259 G; 236 T; 0 other;

Query Match 7.3%; Score 29; DB 23; Length 966;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 CAAAGATCTGTGGGAGAAGTGTGGTTTCC 225
||||||||||||||||||||||||||||
Db 211 CAAAGATCTGTGGGAGAAGTGTGGTTTCC 239

RESULT 4
AAS73441/c
ID AAS73441 standard; cDNA; 660 BP.
XX
AC AAS73441;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9245.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG09254.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 9245; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 660 BP; 149 A; 127 C; 148 G; 236 T; 0 other;

Query Match 6.1%; Score 24; DB 23; Length 660;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 AAAACTCCTGGGTCTCTGTGTGTG 92
||||||||||||||||||||||||
Db 656 AAAACTCCTGGGTCTCTGTGTGTG 633

RESULT 5
AAN60204

```
ID   AAN60204 standard; DNA; 3659 BP.
XX
AC   AAN60204;
XX
DT   05-AUG-1991  (first entry)
XX
DE   Interferon-pseudo-omega-4.
XX
KW   Interferon-omega; virucide; antitumor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2951..3251
FT                   /*tag= a
FT                   /label= interferon-pseudo-omega-4
XX
PN   EP170204-A.
XX
PD   05-FEB-1986.
XX
PF   24-JUL-1985;   85EP-0109284.
XX
PR   14-FEB-1985;   85DE-3505060.
PR   01-AUG-1984;   84DE-3428370.
XX
PA   (BOEH ) BOEHRINGER INGELHEIM.
XX
PI   Hauptmann R,  Meindl P,  Dworkin-Rastl E,  Adolf G,  Swetly P;
PI   Pieler C,  Hauel N;
XX
DR   WPI; 1986-036962/06.
DR   P-PSDB; AAP60256.
XX
PT   New interferon omega polypeptide derivs. - useful as antiviral
PT   and antitumour agents, and new DNA sequences and genes coding for
PT   them.
XX
PS   Disclosure; Fig 14; 115pp; German.
XX
CC   The sequence encodes an interferon analogue, which has virucide
CC   and antitumor activity, and shows a synergistic increase in
CC   activity with interferon-gamma.
XX
SQ   Sequence 3659 BP; 796 A; 885 C; 1039 G; 939 T; 0 other;

  Query Match             5.6%;  Score 22;  DB 7;  Length 3659;
  Best Local Similarity   100.0%;  Pred. No. 0.7;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     194 TTGCAAAGATCTGTGGGAGAAG 215
           ||||||||||||||||||||||
Db    2557 TTGCAAAGATCTGTGGGAGAAG 2578


RESULT 6
ABK83566/c
ID   ABK83566 standard; cDNA; 147724 BP.
XX
AC   ABK83566;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #137.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US30821.
XX
PR   03-OCT-2000; 2000US-237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity -
XX
PS   Claim 1; SEQ ID No 137; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 147724 BP; 46968 A; 29251 C; 28325 G; 43180 T; 0 other;

  Query Match             5.6%;  Score 22;  DB 24;  Length 147724;
  Best Local Similarity   100.0%;  Pred. No. 0.66;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     149 GGCCCTGGTGGCATGGGCTCAC 170
           ||||||||||||||||||||||
Db  123545 GGCCCTGGTGGCATGGGCTCAC 123524


RESULT 7
```

```
AAS91061/c
ID   AAS91061 standard; cDNA; 1179 BP.
XX
AC   AAS91061;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #26865.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG26874.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 26865; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1179 BP; 348 A; 292 C; 315 G; 224 T; 0 other;

  Query Match             5.1%;  Score 20;  DB 23;  Length 1179;
  Best Local Similarity   100.0%;  Pred. No. 6.7;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 GGGTTGCAAAGATCTGTGGG 210
           ||||||||||||||||||||
Db     284 GGGTTGCAAAGATCTGTGGG 265


RESULT 8
AAS66674
ID   AAS66674 standard; cDNA; 2526 BP.
XX
AC   AAS66674;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #2478.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG02487.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 2478; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2526 BP; 871 A; 536 C; 541 G; 578 T; 0 other;

  Query Match             5.1%;  Score 20;  DB 23;  Length 2526;
  Best Local Similarity   100.0%;  Pred. No. 6.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 TTTCATCTTTGCAGGGACTT 51
           ||||||||||||||||||||
Db    2377 TTTCATCTTTGCAGGGACTT 2396


RESULT 9
AAC90702/c
ID   AAC90702 standard; cDNA; 366 BP.
XX
AC   AAC90702;
```

XX
DT 14-MAR-2001 (first entry)
XX
DE Human secretory protein TGC-456 nucleotide sequence SEQ ID NO:17.
XX
KW Human; secretory protein; cancer; immune disease; infectious disease;
KW lung function disorder; liver function disorder; antiinflammatory;
KW gastrointestinal disorder; cytostatic; haematopoietic; anticoagulant;
KW immunomodulatory; hepatotropic; cell proliferation-stimulant;
KW cell migratory agent; cell differentiation-inducer; ss.
XX
OS Homo sapiens.
XX
PN WO200071581-A1.
XX
PD 30-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-JP03221.
XX
20-MAY-1999; 99JP-0140229.
XX
(TAKE) TAKEDA CHEM IND LTD.
XX
PI Itoh Y, Mogi S, Tanaka H, Ohkubo S, Ogi K;
XX
DR WPI; 2001-032023/04.
DR P-PSDB; AAB36662.
XX
PT Novel secretory protein and its salt with e.g. anti-cancer,
PT anti-inflammatory and hematopoietic, effects, applicable as drugs in
PT remedies and preventives to treat diseases like cancer and immune
PT diseases -
XX
PS Example 2; Page 97-98; 122pp; Japanese.
XX
CC AAC90701 to AAC90715 encode the human secretory proteins given in
CC AAB36661 to AAB36675. The proteins can have cytostatic,
CC anti-inflammatory, haematopoietic, anti-coagulant, immunomodulatory and
CC hepatotropic activities, and can be used as cell migratory agents, cell
CC proliferation- stimulants and cell differentiation-inducers. The proteins
CC are useful in the treatment and prevention of diseases such as cancer,
CC lung function disorder, liver function disorder, gastrointestinal
CC disorder and immune diseases. AAC90716 to AAC90755 represent PCR primers
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 366 BP; 99 A; 97 C; 83 G; 87 T; 0 other;

Query Match 4.8%; Score 19; DB 22; Length 366;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 217 GTGGTTTCCTGGATGGGGT 235
||||||||||||||||||||
Db 238 GTGGTTTCCTGGATGGGGT 220

RESULT 10
AAX84949/c
ID AAX84949 standard; DNA; 802 BP.
XX
AC AAX84949;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human secreted protein gene No. 17.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9924836-A1.
XX
PD 20-MAY-1999.
XX
PF 04-NOV-1998; 98WO-US23435.
XX
PR 17-NOV-1997; 97US-0066100.
PR 07-NOV-1997; 97US-0064900.
PR 07-NOV-1997; 97US-0064908.
PR 07-NOV-1997; 97US-0064911.
PR 07-NOV-1997; 97US-0064912.
PR 07-NOV-1997; 97US-0064983.
PR 07-NOV-1997; 97US-0064984.
PR 07-NOV-1997; 97US-0064985.
PR 07-NOV-1997; 97US-0064987.
PR 07-NOV-1997; 97US-0064988.
PR 17-NOV-1997; 97US-0066090.
PR 17-NOV-1997; 97US-0066094.
PR 17-NOV-1997; 97US-0066095.
PR 17-NOV-1997; 97US-0066089.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Ebner R, Endress GA, Feng P, Janat F;
PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y;
XX
DR WPI; 1999-337740/28.
DR P-PSDB; AAY27583.
XX
PT New human secreted proteins and coding sequences useful for treating
PT disorders of the immune system and hyperproliferative disorders
XX
PS Claim 1; Page 290; 507pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number is given in the descriptor line.
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAX84924) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 125 novel genes and their fragments (nucleic
CC acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 125
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX84933 for described uses).
XX
SQ Sequence 802 BP; 208 A; 210 C; 190 G; 188 T; 6 other;

Query Match 4.8%; Score 19; DB 20; Length 802;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 217 GTGGTTTCCTGGATGGGGT 235
||||||||||||||||||||
Db 304 GTGGTTTCCTGGATGGGGT 286

RESULT 11
AAC68089/c
ID AAC68089 standard; cDNA; 1982 BP.
XX
AC AAC68089;
XX
DT 20-FEB-2001 (first entry)
XX

DE Human secreted protein cDNA sequence #9.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200058335-A1.
XX
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US07534.
XX
PR 26-MAR-1999; 99US-0126598.
PR 22-DEC-1999; 99US-0171504.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611702/58.
DR P-PSDB; AAB37356.
XX
PT Nucleic acids encoding human secreted proteins, used to treat, prevent,
PT ameliorate or diagnose conditions such as cancer, and autoimmune
PT diseases e.g. arthritis -
XX
PS Claim 1; Pages 321-322; 387pp; English.
XX
CC The invention relates to the isolation of genes AAC68081-C68127 encoding
CC 47 human secreted proteins AAB37348-B37394. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC68072) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 1982 BP; 549 A; 513 C; 471 G; 448 T; 1 other;

Query Match 4.8%; Score 19; DB 21; Length 1982;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 326 TTTTCTTCATTCTCTGTGG 344
||||||||||||||||||||
Db 860 TTTTCTTCATTCTCTGTGG 842

RESULT 12
ABL09112
ID ABL09112 standard; cDNA; 2770 BP.
XX
AC ABL09112;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21818.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB65009.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 21818; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2770 BP; 786 A; 625 C; 629 G; 730 T; 0 other;

Query Match 4.8%; Score 19; DB 23; Length 2770;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 233 GGTCACACAATCACTCACT 251
||||||||||||||||||||
Db 1050 GGTCACACAATCACTCACT 1068

RESULT 13
ABL28930/c
ID ABL28930 standard; DNA; 4902 BP.
XX
AC ABL28930;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 38263.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 38263; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4902 BP; 1509 A; 1006 C; 984 G; 1403 T; 0 other;

Query Match 4.8%; Score 19; DB 23; Length 4902;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 GGTGGGGCTTTTTTTTGG 286
||||||||||||||||||
Db 886 GGTGGGGCTTTTTTTTGG 868

RESULT 14
AAT45064
ID AAT45064 standard; DNA; 31 BP.
XX
AC AAT45064;
XX
DT 05-FEB-1997 (first entry)
XX
DE 21-hydroxylase fragment A, forward primer.
XX
KW Polymerase chain reaction; primer; PCR; amplify; 21-hydroxylase gene;
KW human adrenal gland cDNA library; diagnosis; detection;
KW Addison's disease; ss.
XX
OS Synthetic.
XX
PN US5376533-A.
XX
PD 27-DEC-1994.
XX
PF 24-MAY-1993; 93US-0066281.
XX
PR 24-MAY-1993; 93US-0066281.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Maclaren NK, Song YH;
XX
DR WPI; 1995-043462/06.
XX
PT Detection of Addison's disease or persons at risk from developing it
PT - using a 21-hydroxylase peptide fragment to detect autoantibodies
PT associated with Addison's disease.
XX
PS Disclosure; Column 6; 9pp; English.
XX
CC The sequences given in AAT45064-75 are primers which were used to
CC amplify fragments of the 21-hydroxylase gene from a human adrenal
CC gland cDNA library. The amplified fragments encode peptides which
CC were used in the method of the invention to diagnose or detect
CC Addison's disease. Fragment A comprises amino acids 1-162, fragment
CC B, amino acids 164-356, fragment C, amino acids 3440494, fragment D,
CC amino acids 164-271, fragment E, amino acids 272-356, and fragment F
CC amino acids 197-298 of the 21-hydroxylase enzyme. These primers
CC have been optimised for use and do not directly correspond to 21
CC hydroxylase gene sequences. All the forward primers contain BamHI
CC sites and all the reverse primers contain EcoRI restriction sites.
XX
SQ Sequence 31 BP; 2 A; 10 C; 11 G; 8 T; 0 other;

Query Match 4.5%; Score 18; DB 16; Length 31;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 288 TCCATGCTGCTCCTGGGC 305
||||||||||||||||||
Db 8 TCCATGCTGCTCCTGGGC 25

RESULT 15
AAZ43080/c
ID AAZ43080 standard; DNA; 339 BP.
XX
AC AAZ43080;
XX
DT 04-FEB-2000 (first entry)
XX
DE C. elegans insulin-like protein ZK84.6 gene.
XX
KW Insulin-like protein; diagnosis; insulin-like gene analysis; nematode;
KW insulin hormone; aging; senescence; pesticide target; signalling pathway;
KW signal transduction pathway; ss.
XX
OS Caenorhabditis elegans.
XX
PN WO9954436-A2.
XX
PD 28-OCT-1999.
XX
PF 16-APR-1999; 99WO-US08522.
XX
PR 17-APR-1998; 98US-0062580.
PR 08-MAY-1998; 98US-0074984.
PR 26-MAY-1998; 98US-0084303.
XX
PA (EXEL-) EXELIXIS PHARM INC.
XX
PI Homburger SA, Platt DM, Ferguson KC, Doberstein SK, Buchman AR;
PI Reddy BP;
XX
DR WPI; 2000-013239/01.
DR P-PSDB; AAY65657.
XX
PT Analysing Caenorhabditis elegans insulin-like gene expression, nucleic
PT acids and proteins of the C. elegans insulin-like genes -
XX
PS Claim 4; Fig 8; 194pp; English.
XX
CC This sequence encodes a Caenorhabditis elegans insulin-like protein,
CC and can be used in the method of the invention. The method is for
CC analysing an effect of expression or mis-expression of a C. elegans
CC insulin-like gene, and comprises observing a first nematode genetically
CC engineered to express or mis-express a C. elegans insulin-like protein
CC (ILP) of any one of groups I, II or IV or a derivative or fragment that
CC displays one or more functional activities of the C. elegans ILP. The
CC insulin-like genes in C. elegans constitute very useful tools for probing
CC the function and regulation of their corresponding pathways. This can be
CC expected to lead to the discovery of new drug targets, therapeutic
CC proteins, diagnostics and prognostics useful in the treatment of diseases

CC and clinical problems associated with the function of insulin hormones in
CC humans and other animals, as well as clinical problems associated with
CC aging and senescence. The information may also be useful in
CC identification and validation of pesticide targets in invertebrate pests
CC that are components of these signalling pathways. The genes are also
CC useful for identifying factors that are upstream of the receptor in the
CC signal transduction pathway. The ligand-encoding C. elegans insulin-like
CC genes provide a superior approach for identifying factors that are
CC upstream of the receptor in the signal transduction pathway.
XX
SQ Sequence 339 BP; 92 A; 86 C; 72 G; 89 T; 0 other;

Query Match 4.5%; Score 18; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 188 TGTGGGTTGCAAAGATCT 205
||||||||||||||||||
Db 260 TGTGGGTTGCAAAGATCT 243

arch completed: April 25, 2003, 00:01:03
Job time : 220.988 secs

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:49:04 ; Search time 1331.72 Seconds
(without alignments)
4815.901 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_446
Perfect score: 396
Sequence: 1 atgggtggatcttttgcctt.....gamctggatatttcagttga 396

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

d size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	147	37.1	470	17	AQ770688	AQ770688 HS_5368_B
	2	146	36.9	525	17	AQ165256	AQ165256 HS_3025_B
	3	38	9.6	628	17	AQ237815	AQ237815 RPCI11-70
	4	35	8.8	410	17	AQ442274	AQ442274 HS_5137_A
	5	35	8.8	453	17	AQ437684	AQ437684 HS_5137_A
c	6	35	8.8	635	17	AQ390599	AQ390599 CITBI-E1-
	7	34	8.6	412	17	AQ321916	AQ321916 RPCI11-10
c	8	34	8.6	529	17	AQ881246	AQ881246 HS_5137_B
c	9	33	8.3	378	17	AQ479650	AQ479650 RPCI-11-2
c	10	33	8.3	471	17	AQ147593	AQ147593 HS_3065_B
c	11	33	8.3	482	17	AQ320567	AQ320567 RPCI11-99
c	12	33	8.3	546	12	BF962702	BF962702 PM4-NN120
c	13	33	8.3	563	17	AQ420187	AQ420187 RPCI-11-1
c	14	33	8.3	691	17	AZ516454	AZ516454 RPCI-11-1
c	15	33	8.3	723	17	AQ386439	AQ386439 RPCI11-15
	16	32	8.1	427	17	AQ697116	AQ697116 HS_5528_A
	17	31	7.8	553	17	AZ521751	AZ521751 RPCI-11-1
	18	31	7.8	695	17	AG179297	AG179297 Pan trogl
	19	31	7.8	769	17	AQ899390	AQ899390 HS_5234_A
	20	30	7.6	360	17	AQ207172	AQ207172 HS_3239_B
c	21	30	7.6	399	17	AQ115544	AQ115544 RPCI11-57
c	22	30	7.6	435	17	AQ116061	AQ116061 RPCI11-57
c	23	30	7.6	452	17	AQ442744	AQ442744 HS_5122_B
	24	30	7.6	551	17	AQ569689	AQ569689 HS_5333_B
	25	29	7.3	414	17	AQ003326	AQ003326 RPCI11-25
c	26	29	7.3	468	17	AQ819715	AQ819715 HS_5513_A
c	27	29	7.3	615	17	AG161224	AG161224 Pan trogl
c	28	29	7.3	653	17	AG143347	AG143347 Pan trogl
c	29	28	7.1	363	17	AQ120796	AQ120796 HS_3076_A
	30	28	7.1	376	17	AQ548294	AQ548294 RPCI-11-4
	31	28	7.1	401	17	AQ588089	AQ588089 CITBI-E1-
	32	28	7.1	419	17	AQ436209	AQ436209 HS_5049_B
	33	28	7.1	425	17	AQ141070	AQ141070 HS_3141_B
c	34	28	7.1	449	17	AQ269217	AQ269217 RPCI11-69
c	35	28	7.1	490	17	AQ817757	AQ817757 HS_5250_B
c	36	28	7.1	530	17	AQ193128	AQ193128 HS_3060_B
c	37	28	7.1	541	17	AQ683783	AQ683783 HS_5455_B
	38	28	7.1	653	17	AG160919	AG160919 Pan trogl
c	39	28	7.1	681	17	AG141287	AG141287 Pan trogl
	40	28	7.1	684	17	AG049745	AG049745 Pan trogl
	41	27	6.8	345	17	AQ075574	AQ075574 CIT-HSP-2
c	42	27	6.8	517	17	AQ614252	AQ614252 HS_5123_B
	43	27	6.8	529	17	AQ346593	AQ346593 RPCI11-12
	44	27	6.8	553	17	AQ238365	AQ238365 RPCI11-63
	45	27	6.8	606	17	AQ350708	AQ350708 RPCI11-11

ALIGNMENTS

RESULT 1
AQ770688
LOCUS AQ770688 470 bp DNA linear GSS 28-JUL-1999
DEFINITION HS_5368_B2_C08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=944 Col=16 Row=F, DNA sequence.
ACCESSION AQ770688
VERSION AQ770688.1 GI:5648804
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 944  row: F  column: 16
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=944 Col=16 Row=F"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
[redacted]E COUNT       83 a    112 c    131 g    141 t      3 others
ORIGIN

  Query Match             37.1%;  Score 147;  DB 17;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-54;
  Matches  147;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 162

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTG 222

Qy     121 CATACACAGCTCTGTATATCGGACCCA 147
           ||||||||||||||||||||||||||
Db     223 CATACACAGCTCTGTATATCGGACCCA 249


RESULT 2
AQ165256
LOCUS       AQ165256                525 bp    DNA     linear   GSS 16-OCT-1998
DEFINITION  HS_3025_B2_G06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=12 Row=N, DNA sequence.
ACCESSION   AQ165256
VERSION     AQ165256.1  GI:3563451
KEYWORDS    GSS.
[redacted]RCE       human.
[redacted]RGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3025  row: N  column: 12
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES              Location/Qualifiers
     source           1. .525
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3025 Col=12 Row=N"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      102 a    139 c    137 g    143 t      4 others
ORIGIN

  Query Match             36.9%;  Score 146;  DB 17;  Length 525;
  Best Local Similarity   99.5%;  Pred. No. 1.1e-53;
  Matches  196;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      80 GTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTGCATACACAGCTCTGTATAT 139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 GTCTCTGTGTGTGCCTGAGTGGCTGCTCTACTGAGACTCTGCATACACAGCTCTGTATAT 204

Qy     140 CGGACCCAWGGCCCTGGTGGCATGGGCTCACGAGGAGATCCCCTGATCTGTGGGTTGCAA 199
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 CGGACCCAAGGCCCTGGTGGCATGGGCTCACGAGGAGATCCCCTGATCTGTGGGTTGCAA 264

Qy     200 AGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACACAATCACTCACTGCTTCCCT 259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 AGATCTGTGGGAGAAGTGTGGTTTCCTGGATGGGGTCACACAATCACTCACTGCTTCCCT 324

Qy     260 TGGCTGGAGGTGGGGCT 276
           |||||||||||||||||
Db     325 TGGCTGGAGGTGGGGCT 341


RESULT 3
AQ237815
LOCUS       AQ237815                628 bp    DNA     linear   GSS 21-APR-1999
DEFINITION  RPCI11-70H4.TK RPCI-11 Homo sapiens genomic clone RPCI-11-70H4, DNA
            sequence.
ACCESSION   AQ237815
VERSION     AQ237815.1  GI:3670106
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .628
                      /organism="Homo sapiens"
                      /db_xref="GDB:7526667"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-70H4"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
```

RPCI11 Human Male BAC Library"
BASE COUNT 125 a 150 c 164 g 189 t
ORIGIN

Query Match 9.6%; Score 38; DB 17; Length 628;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
||||||||||||||||||||||||||||||||||||||
Db 157 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 194

RESULT 4
AQ442274
LOCUS AQ442274 410 bp DNA linear GSS 31-MAR-1999
DEFINITION HS_5137_A1_F12_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=713 Col=23 Row=K, DNA sequence.
ACCESSION AQ442274
VERSION AQ442274.1 GI:4553613
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 410)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 713 row: K column: 23
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 410.
FEATURES Location/Qualifiers
source 1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=713 Col=23 Row=K"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 80 a 111 c 111 g 107 t 1 others
ORIGIN

Query Match 8.8%; Score 35; DB 17; Length 410;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
|||||||||||||||||||||||||||||||||||
Db 160 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 194

RESULT 5
AQ437684
LOCUS AQ437684 453 bp DNA linear GSS 31-MAR-1999
DEFINITION HS_5137_A2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=713 Col=12 Row=O, DNA sequence.
ACCESSION AQ437684
VERSION AQ437684.1 GI:4549023
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 713 row: O column: 12
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 453.
FEATURES Location/Qualifiers
source 1. .453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=713 Col=12 Row=O"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 84 a 127 c 117 g 124 t 1 others
ORIGIN

Query Match 8.8%; Score 35; DB 17; Length 453;
Best Local Similarity 100.0%; Pred. No. 5.8e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
|||||||||||||||||||||||||||||||||||
Db 163 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 197

RESULT 6
AQ390599/c
LOCUS AQ390599 635 bp DNA linear GSS 06-MAR-1999
DEFINITION CITBI-E1-2544B15.TR CITBI-E1 Homo sapiens genomic clone 2544B15, DNA sequence.
ACCESSION AQ390599
VERSION AQ390599.1 GI:4361622
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2544B15.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
    URES            Location/Qualifiers
     source         1. .635
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="2544B15"
                    /clone_lib="CITBI-E1"
                    /sex="male"
                    /cell_type="sperm"
                    /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                    CalTech Human BAC Library D"
BASE COUNT      192 a    144 c    172 g    127 t
ORIGIN

  Query Match             8.8%;  Score 35;  DB 17;  Length 635;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-05;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 108
           |||||||||||||||||||||||||||||||||||
Db     527 TCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCT 493


RESULT 7
AQ321916
LOCUS       AQ321916                412 bp    DNA     linear   GSS 06-MAY-1999
DEFINITION  RPCI11-101H18.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-101H18,
            DNA sequence.
ACCESSION   AQ321916
VERSION     AQ321916.1  GI:4054584
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-101H18.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .412
                    /organism="Homo sapiens"
                    /db_xref="GDB:7538585"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-101H18"
                    /clone_lib="RPCI-11"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
BASE COUNT      74 a    102 c    120 g    116 t
ORIGIN

  Query Match             8.6%;  Score 34;  DB 17;  Length 412;
  Best Local Similarity   100.0%;  Pred. No. 0.00016;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 GGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTC 224
           ||||||||||||||||||||||||||||||||||
Db     292 GGGTTGCAAAGATCTGTGGGAGAAGTGTGGTTTC 325


RESULT 8
AQ881246/c
LOCUS       AQ881246                529 bp    DNA     linear   GSS 09-NOV-1999
DEFINITION  HS_5137_B1_F08_T7 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=8905 Col=15 Row=L, DNA sequence.
ACCESSION   AQ881246
VERSION     AQ881246.1  GI:6312713
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 8905  row: L  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 529.
FEATURES            Location/Qualifiers
     source         1. .529
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate=8905 Col=15 Row=L"
                    /clone_lib="RPCI-11 Human Male BAC Library"
                    /sex="male"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    Male blood DNA was isolated from one randomly chosen donor
                    and partially digested with a combination of EcoRI and
                    EcoRI Methylase. Size selected DNA was cloned into the
```

pBACe3.6 vector at EcoRI sites"
BASE COUNT 137 a 147 c 121 g 117 t 7 others
ORIGIN

Query Match 8.6%; Score 34; DB 17; Length 529;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGC 93
||||||||||||||||||||||||||||||||||
Db 382 GGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGC 349

RESULT 9
AQ479650/c
LOCUS AQ479650 378 bp DNA linear GSS 23-APR-1999
DEFINITION RPCI-11-269D19.TV RPCI-11 Homo sapiens genomic clone RPCI-11-269D19 , DNA sequence.
ACCESSION AQ479650
VERSION AQ479650.1 GI:4661769
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 378)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter ,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .378
/organism="Homo sapiens"
/db_xref="GDB:7603002"
/db_xref="taxon:9606"
/clone="RPCI-11-269D19"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 116 a 103 c 93 g 66 t
ORIGIN

Query Match 8.3%; Score 33; DB 17; Length 378;
Best Local Similarity 100.0%; Pred. No. 0.00046;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 225
|||||||||||||||||||||||||||||||||
Db 208 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 176

RESULT 10
AQ147593/c
LOCUS AQ147593 471 bp DNA linear GSS 08-OCT-1998
DEFINITION HS_3065_B2_H08_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3065 Col=16 Row=P, DNA sequence.
ACCESSION AQ147593
VERSION AQ147593.1 GI:3538246
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3065 row: P Column: 16
Class: BAC ends
High quality sequence stop: 471.
FEATURES Location/Qualifiers
source 1. .471
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3065 Col=16 Row=P"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 144 a 118 c 123 g 84 t 2 others
ORIGIN

Query Match 8.3%; Score 33; DB 17; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 225
|||||||||||||||||||||||||||||||||
Db 265 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 233

RESULT 11
AQ320567/c
LOCUS AQ320567 482 bp DNA linear GSS 04-MAY-1999
DEFINITION RPCI11-99N1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-99N1, DNA sequence.
ACCESSION AQ320567
VERSION AQ320567.1 GI:4050696
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Homo sapiens"
                     /db_xref="GDB:7537944"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-99N1"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      142 a    121 c    106 g    112 t      1 others
ORIGIN

  Query Match             8.3%;  Score 33;  DB 17;  Length 482;
  Best Local Similarity   100.0%;  Pred. No. 0.00044;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
Db     338 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 306


RESULT 12
BF962702/c
LOCUS       BF962702                546 bp    mRNA    linear   EST 22-JAN-2001
DEFINITION  PM4-NN1209-151200-012-f04 NN1209 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF962702
VERSION     BF962702.1  GI:12379977
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
   ITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-NN1209-
            151200-012-f04&t3=2000-12-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN1209"
                     /dev_stage="Adult"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      149 a    157 c    135 g    105 t
ORIGIN

  Query Match             8.3%;  Score 33;  DB 12;  Length 546;
  Best Local Similarity   100.0%;  Pred. No. 0.00042;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     193 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 225
           |||||||||||||||||||||||||||||||||
Db      96 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 64


RESULT 13
AQ420187/c
LOCUS       AQ420187                563 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-185J19.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-185J19
            , DNA sequence.
ACCESSION   AQ420187
VERSION     AQ420187.1  GI:4477911
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-185J19.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="GDB:7570890"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-185J19"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      173 a    127 c    114 g    149 t
ORIGIN

  Query Match             8.3%;  Score 33;  DB 17;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 0.00042;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
```

```
Db    333 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 301


RESULT 14
AZ516454/c
LOCUS       AZ516454                691 bp    DNA     linear   GSS 16-OCT-2000
DEFINITION  RPCI-11-191D19.TV RPCI-11 Homo sapiens genomic clone RPCI-11-191D19
            , DNA sequence.
ACCESSION   AZ516454
VERSION     AZ516454.1  GI:10824878
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     BAC end sequences of library RPCI-11
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-191D19.TJB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            This BAC end was generated during the R&D process and may have
            higher chance of clone tracking errors.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Homo sapiens"
                     /db_xref="GDB:7573050"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-191D19"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      203 a    191 c    174 g    121 t      2 others
ORIGIN

  Query Match             8.3%;  Score 33;  DB 17;  Length 691;
  Best Local Similarity   100.0%;  Pred. No. 0.0004;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     193 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 225
           |||||||||||||||||||||||||||||||||
Db     216 GTTGCAAAGATCTGTGGGAGAAGTGTGGTTTCC 184


RESULT 15
AQ386439/c
LOCUS       AQ386439                723 bp    DNA     linear   GSS 21-MAY-1999
DEFINITION  RPCI11-154D6.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-154D6,
            DNA sequence.
ACCESSION   AQ386439
VERSION     AQ386439.1  GI:4357462
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-154D6.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Homo sapiens"
                     /db_xref="GDB:7558829"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-154D6"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      224 a    165 c    135 g    199 t
ORIGIN

  Query Match             8.3%;  Score 33;  DB 17;  Length 723;
  Best Local Similarity   100.0%;  Pred. No. 0.0004;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 85
           |||||||||||||||||||||||||||||||||
Db     331 TGGGGCCGGAGTATGTAAAACTCCTGGGTCTCT 299


Search completed: April 25, 2003, 00:52:51
Job time : 1334.72 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 23:57:29 ; Search time 46.8639 Seconds
(without alignments)
2591.419 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_446
Perfect score: 396
Sequence: 1 atgggtggatcttttgcctt.....gamctggatatttcagttga 396

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

d size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	4.5	31	1	US-08-066-281-1	Sequence 1, Appli
c	2	18	4.5	430	4	US-09-397-787-254	Sequence 254, App
	3	17	4.3	1118	4	US-09-452-239-37	Sequence 37, Appl
	4	17	4.3	1146	4	US-09-452-239-3	Sequence 3, Appli
c	5	17	4.3	1381	4	US-08-858-207A-108	Sequence 108, App
	6	17	4.3	1509	1	US-08-115-052-1	Sequence 1, Appli
	7	17	4.3	2686	4	US-09-228-986-3	Sequence 3, Appli
c	8	17	4.3	11613	4	US-09-453-702B-42	Sequence 42, Appl
	9	17	4.3	40328	3	US-08-742-185-102	Sequence 102, App
c	10	16	4.0	36	4	US-08-910-722-7	Sequence 7, Appli
c	11	16	4.0	42	4	US-08-910-722-5	Sequence 5, Appli
c	12	16	4.0	57	1	US-08-474-177-17	Sequence 17, Appl
c	13	16	4.0	57	1	US-08-487-033-17	Sequence 17, Appl
c	14	16	4.0	57	1	US-08-480-810-17	Sequence 17, Appl
c	15	16	4.0	57	2	US-08-508-735-17	Sequence 17, Appl
c	16	16	4.0	57	2	US-08-848-251-17	Sequence 17, Appl
c	17	16	4.0	57	2	US-08-486-047-17	Sequence 17, Appl
c	18	16	4.0	57	3	US-09-120-130-17	Sequence 17, Appl
c	19	16	4.0	57	3	US-09-115-252-17	Sequence 17, Appl
c	20	16	4.0	57	3	US-08-986-515-17	Sequence 17, Appl
c	21	16	4.0	57	3	US-09-120-128-17	Sequence 17, Appl
c	22	16	4.0	57	4	US-09-120-129-17	Sequence 17, Appl
c	23	16	4.0	57	4	US-09-201-139-17	Sequence 17, Appl
c	24	16	4.0	57	4	US-09-120-131-17	Sequence 17, Appl
	25	16	4.0	384	4	US-09-134-001C-1752	Sequence 1752, Ap
c	26	16	4.0	471	1	US-08-474-177-1	Sequence 1, Appli
c	27	16	4.0	471	1	US-08-487-033-1	Sequence 1, Appli
c	28	16	4.0	471	1	US-08-480-810-1	Sequence 1, Appli
c	29	16	4.0	471	2	US-08-508-735-1	Sequence 1, Appli
c	30	16	4.0	471	2	US-08-848-251-1	Sequence 1, Appli
c	31	16	4.0	471	2	US-08-486-047-1	Sequence 1, Appli
c	32	16	4.0	471	3	US-09-120-130-1	Sequence 1, Appli
c	33	16	4.0	471	3	US-09-115-252-1	Sequence 1, Appli
c	34	16	4.0	471	3	US-08-986-515-1	Sequence 1, Appli
c	35	16	4.0	471	3	US-09-120-128-1	Sequence 1, Appli
c	36	16	4.0	471	4	US-09-120-129-1	Sequence 1, Appli
c	37	16	4.0	471	4	US-09-201-139-1	Sequence 1, Appli
c	38	16	4.0	471	4	US-09-120-131-1	Sequence 1, Appli
c	39	16	4.0	536	4	US-09-221-017B-1081	Sequence 1081, Ap
c	40	16	4.0	687	4	US-09-457-568-23	Sequence 23, Appl
c	41	16	4.0	687	4	US-09-457-646-23	Sequence 23, Appl
c	42	16	4.0	737	4	US-09-457-568-19	Sequence 19, Appl
c	43	16	4.0	737	4	US-09-457-646-19	Sequence 19, Appl
c	44	16	4.0	782	4	US-09-457-568-21	Sequence 21, Appl
c	45	16	4.0	782	4	US-09-457-646-21	Sequence 21, Appl

ALIGNMENTS

```
RESULT 1
US-08-066-281-1
; Sequence 1, Application US/08066281
; Patent No. 5376533
;  GENERAL INFORMATION:
;    APPLICANT:  Maclaren, No. 53765331 K.
;    APPLICANT:  Song, Yao Hua
;    TITLE OF INVENTION:  Methods and Compositions for the Detection of
;    TITLE OF INVENTION:  Addison's Disease
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID R. SALIWANCHIK
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/066,281
;      FILING DATE:  19930524
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF/S&S-132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (synthetic)
US-08-066-281-1

  Query Match             4.5%;  Score 18;  DB 1;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     288 TCCATGCTGCTCCTGGGC 305
           ||||||||||||||||||
Db       8 TCCATGCTGCTCCTGGGC 25
```

```
RESULT 2
US-09-397-787-254/c
; Sequence 254, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 254
;   LENGTH: 430
    TYPE: DNA
    ORGANISM: Homo sapien
    FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(430)
;   OTHER INFORMATION: n = A,T,C or G
US-09-397-787-254

  Query Match             4.5%;  Score 18;  DB 4;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    191 GGGTTGCAAAGATCTGTG 208
          ||||||||||||||||||
Db    273 GGGTTGCAAAGATCTGTG 256


RESULT 3
US-09-452-239-37
; Sequence 37, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
;  APPLICANT: Rafalski, Antoni J.
;  APPLICANT: Fader, Gary M.
;  APPLICANT: Cahoon, Rebecca E.
;  TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
;  FILE REFERENCE: BB1284 US NA
;  CURRENT APPLICATION NUMBER: US/09/452,239
;  CURRENT FILING DATE: 1999-12-01
   EARLIER APPLICATION NUMBER: 60/110,594
   EARLIER FILING DATE: 1998-December-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
;   LENGTH: 1118
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-452-239-37

  Query Match             4.3%;  Score 17;  DB 4;  Length 1118;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    273 GGCTTTTTTTTTGGCTC 289
          |||||||||||||||||
Db    971 GGCTTTTTTTTTGGCTC 987


RESULT 4
US-09-452-239-3
; Sequence 3, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
;  APPLICANT: Rafalski, Antoni J.
;  APPLICANT: Fader, Gary M.
;  APPLICANT: Cahoon, Rebecca E.
;  TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
;  FILE REFERENCE: BB1284 US NA
;  CURRENT APPLICATION NUMBER: US/09/452,239
;  CURRENT FILING DATE: 1999-12-01
;  EARLIER APPLICATION NUMBER: 60/110,594
;  EARLIER FILING DATE: 1998-December-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 1146
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-452-239-3

  Query Match             4.3%;  Score 17;  DB 4;  Length 1146;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    273 GGCTTTTTTTTTGGCTC 289
          |||||||||||||||||
Db    976 GGCTTTTTTTTTGGCTC 992


RESULT 5
US-08-858-207A-108/c
; Sequence 108, Application US/08858207A
; Patent No. 6348328
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Stodola, Robert
;    TITLE OF INVENTION:  No. 6348328el Compounds
;    NUMBER OF SEQUENCES:  552
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,207A
;      FILING DATE:  09-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017670
;      FILING DATE:  14-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  108:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-858-207A-108

  Query Match             4.3%;  Score 17;  DB 4;  Length 1381;
```

```
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     323 CATTTTTCTTCATTCTC 339
           |||||||||||||||||
Db     295 CATTTTTCTTCATTCTC 279


RESULT 6
US-08-115-052-1
; Sequence 1, Application US/08115052
; Patent No. 5705400
;  GENERAL INFORMATION:
;    APPLICANT:  Furmaniak-Wehr, Jadwiga Maria
;    TITLE OF INVENTION:  Assay for Adrenal Autoantigen
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street N.W.
       CITY:  Washington
       STATE:  D.C.
       COUNTRY:  USA
;      ZIP:          20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/115,052
;      FILING DATE:  02-SEP-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/937,409
;      FILING DATE:  31-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dale H. Hoscheit
;      REGISTRATION NUMBER:  19,090
;      REFERENCE/DOCKET NUMBER:  01950.44179
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-5089299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1509 base pairs
;      TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
     MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      DEVELOPMENTAL STAGE:  foetus
;      TISSUE TYPE:  adrenal gland
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  13..54
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  55..1494
;      OTHER INFORMATION:  /product= "steroid 21-hydroxylase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  13..1494
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (435^436)
;      OTHER INFORMATION:  /standard_name= "PvuII cleavage
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (732^733)
;      OTHER INFORMATION:  /standard_name= "PvuII cleavage
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (852^853)
;      OTHER INFORMATION:  /standard_name= "PmaCI cleavage
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (1148^1149)
;      OTHER INFORMATION:  /standard_name= "SauI cleavage
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (1354^1355)
;      OTHER INFORMATION:  /standard_name= "StuI cleavage
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  (53^54)
;      OTHER INFORMATION:  /standard_name= "NarI cleavage
;      OTHER INFORMATION:  site"
US-08-115-052-1

  Query Match             4.3%;  Score 17;  DB 1;  Length 1509;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     289 CCATGCTGCTCCTGGGC 305
           |||||||||||||||||
Db      11 CCATGCTGCTCCTGGGC 27


RESULT 7
US-09-228-986-3
; Sequence 3, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
;  APPLICANT: Strabala, Timothy
;  APPLICANT: Nieuwenhuizen, Niels
;  TITLE OF INVENTION: Compositions Isolated from Plant Cells
;  TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
;  FILE REFERENCE: 11000/1020
;  CURRENT APPLICATION NUMBER: US/09/228,986
;  CURRENT FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 130
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2686
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-228-986-3

  Query Match             4.3%;  Score 17;  DB 4;  Length 2686;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     318 TACCCCATTTTTCTTCA 334
           |||||||||||||||||
Db     272 TACCCCATTTTTCTTCA 288


RESULT 8
US-09-453-702B-42/c
; Sequence 42, Application US/09453702B
; Patent No. 6365723
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
```

```
;       TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;       NUMBER OF SEQUENCES: 265
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Quarles & Brady
;            STREET: 1 South Pinckney Street
;            CITY: Madison
;            STATE: WI
;            COUNTRY: US
;            ZIP: 53701-2113
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 8.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/453,702B
;            FILING DATE: 03-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/110,955
;            FILING DATE: 04-DEC-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seay, Nicholas J.
;            REGISTRATION NUMBER: 27386
;            REFERENCE/DOCKET NUMBER: 960296.95017
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (608) 251-5000
;            TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 42:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 11613
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA  (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-453-702B-42

  Query Match             4.3%;  Score 17;  DB 4;  Length 11613;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     219 GGTTTCCTGGATGGGGT 235
           |||||||||||||||||
Db    2106 GGTTTCCTGGATGGGGT 2090


RESULT 9
US-08-742-185-102
; Sequence 102, Application US/08742185
; Patent No. 6020476
;  GENERAL INFORMATION:
;    APPLICANT:  Page, David C.
;    APPLICANT:  Reijo, Renee
;    APPLICANT:  Saxena, Richa
;    APPLICANT:  Hawkins,Trevor
;    APPLICANT:  Reeve, Mary Pat
;    TITLE OF INVENTION:  DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,185
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/690,734
;      FILING DATE:  31-JUL-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/310,429
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI94-07A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  102:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-742-185-102

  Query Match             4.3%;  Score 17;  DB 3;  Length 40328;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     268 GGTGGGGCTTTTTTTTT 284
           |||||||||||||||||
Db     882 GGTGGGGCTTTTTTTTT 898


RESULT 10
US-08-910-722-7/c
; Sequence 7, Application US/08910722
; Patent No. 6251871
;  GENERAL INFORMATION:
;    APPLICANT:  Jin, Xiaomei
;    APPLICANT:  Roth, Jack A.
;    TITLE OF INVENTION:  p16 EXPRESSION CONSTRUCTS AND THEIR
;    TITLE OF INVENTION:  APPLICATION IN CANCER THERAPY
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,722
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/502,881
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  INGN:016/HYL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  36 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-910-722-7

  Query Match             4.0%;  Score 16;  DB 4;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    285 GGCTCCATGCTGCTCC 300
          ||||||||||||||||
Db     23 GGCTCCATGCTGCTCC 8


RESULT 11
US-08-910-722-5/c
; Sequence 5, Application US/08910722
; Patent No. 6251871
;  GENERAL INFORMATION:
;    APPLICANT:  Jin, Xiaomei
;    APPLICANT:  Roth, Jack A.
;    TITLE OF INVENTION:  p16 EXPRESSION CONSTRUCTS AND THEIR
;    TITLE OF INVENTION:  APPLICATION IN CANCER THERAPY
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,722
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/502,881
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  INGN:016/HYL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  42 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-910-722-5

  Query Match             4.0%;  Score 16;  DB 4;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    285 GGCTCCATGCTGCTCC 300
          ||||||||||||||||
Db     29 GGCTCCATGCTGCTCC 14


RESULT 12
```

```
US-08-474-177-17/c
; Sequence 17, Application US/08474177
; Patent No. 5624819
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Cannon-Albright, Lisa A.
;    APPLICANT:  Kamb, Alexander
;    TITLE OF INVENTION:  GERMLINE MUTATIONS IN THE MTS GENE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,177
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03537
;      FILING DATE:  17-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/251,938
;      FILING DATE:  01-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,087
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,086
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/227,369
;      FILING DATE:  14-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/214,582
;      FILING DATE:  18-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109348-E
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  57 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-474-177-17

  Query Match             4.0%;  Score 16;  DB 1;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    285 GGCTCCATGCTGCTCC 300
          ||||||||||||||||
Db     50 GGCTCCATGCTGCTCC 35


RESULT 13
```

```
US-08-487-033-17/c
; Sequence 17, Application US/08487033
; Patent No. 5739027
;  GENERAL INFORMATION:
;    APPLICANT:  Kamb, Alexander
;    TITLE OF INVENTION:  MTS1E1-Beta GENE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,033
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03316
;      FILING DATE:  17-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/251,938
;      FILING DATE:  01-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,087
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,086
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/227,369
;      FILING DATE:  14-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/214,582
;      FILING DATE:  18-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109348-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
   INFORMATION FOR SEQ ID NO:  17:
    SEQUENCE CHARACTERISTICS:
;     LENGTH:  57 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-487-033-17

  Query Match             4.0%;  Score 16;  DB 1;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     285 GGCTCCATGCTGCTCC 300
           ||||||||||||||||
Db      50 GGCTCCATGCTGCTCC 35


RESULT 14
US-08-480-810-17/c
; Sequence 17, Application US/08480810
; Patent No. 5801236
;  GENERAL INFORMATION:
;    APPLICANT:  Kamb, Alexander
;    TITLE OF INVENTION:  MTS1 GENE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,810
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03316
;      FILING DATE:  17-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/251,938
;      FILING DATE:  01-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,087
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/215,086
;      FILING DATE:  18-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/227,369
;      FILING DATE:  14-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/214,582
;      FILING DATE:  18-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109348
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  57 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-480-810-17

  Query Match             4.0%;  Score 16;  DB 1;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     285 GGCTCCATGCTGCTCC 300
           ||||||||||||||||
Db      50 GGCTCCATGCTGCTCC 35


RESULT 15
US-08-508-735-17/c
; Sequence 17, Application US/08508735
; Patent No. 5843756
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Stone, Steven
;    APPLICANT:  Jiang, Ping
;    APPLICANT:  Kamb, Alexander
;    TITLE OF INVENTION:  MTS GENE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,735
       FILING DATE:
       CLASSIFICATION:  435
     PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US to be assigned
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/03316
;      FILING DATE:  17-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109348
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4848
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  57 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-508-735-17

  Query Match             4.0%;  Score 16;  DB 2;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     285 GGCTCCATGCTGCTCC 300
           ||||||||||||||||
Db      50 GGCTCCATGCTGCTCC 35


Search completed: April 25, 2003, 00:54:16
Job time : 56.8639 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 00:41:49 ; Search time 81.2308 Seconds
(without alignments)
5304.628 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_446
Perfect score: 396
Sequence: 1 atgggtggatcttttgcctt.....gamctggatatttcagttga 396

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 709820 seqs, 544064369 residues

d size : 0

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	33	8.3	684973	10	US-09-263-959-1	Sequence 1, Appli
c	2	22	5.6	361	9	US-09-918-995-8549	Sequence 8549, Ap
c	3	19	4.8	802	9	US-09-974-879-27	Sequence 27, Appl
c	4	18	4.5	430	10	US-09-876-889-254	Sequence 254, App
	5	18	4.5	431	10	US-09-866-562-80	Sequence 80, Appl
	6	18	4.5	431	10	US-09-866-562-87	Sequence 87, Appl
c	7	18	4.5	570	10	US-09-864-761-9118	Sequence 9118, Ap
c	8	17	4.3	489	9	US-09-918-995-28049	Sequence 28049, A
c	9	17	4.3	491	9	US-09-918-995-34650	Sequence 34650, A
c	10	17	4.3	492	9	US-09-918-995-19262	Sequence 19262, A
c	11	17	4.3	531	9	US-10-092-154-1899	Sequence 1899, Ap
c	12	17	4.3	531	10	US-09-764-847-1899	Sequence 1899, Ap
c	13	17	4.3	553	10	US-09-864-761-13668	Sequence 13668, A
c	14	17	4.3	609	10	US-09-974-300-8084	Sequence 8084, Ap
c	15	17	4.3	944	9	US-09-774-639-99	Sequence 99, Appl
c	16	17	4.3	944	9	US-09-969-730-16	Sequence 16, Appl
	17	17	4.3	1047	10	US-09-822-830A-468	Sequence 468, App
c	18	17	4.3	1089	10	US-09-962-740-3	Sequence 3, Appli
	19	17	4.3	1118	10	US-09-452-239-37	Sequence 37, Appl
	20	17	4.3	1146	10	US-09-452-239-3	Sequence 3, Appli
c	21	17	4.3	1215	10	US-09-962-740-1	Sequence 1, Appli
c	22	17	4.3	1215	10	US-09-962-740-6	Sequence 6, Appli
c	23	17	4.3	1396	10	US-09-962-740-8	Sequence 8, Appli
c	24	17	4.3	1631	10	US-09-962-740-5	Sequence 5, Appli
	25	17	4.3	2686	9	US-10-101-464A-3	Sequence 3, Appli
	26	17	4.3	2880	9	US-09-951-502A-1	Sequence 1, Appli
c	27	17	4.3	11613	9	US-10-114-170-42	Sequence 42, Appl
	28	17	4.3	38374	10	US-09-880-107-3463	Sequence 3463, Ap
	29	17	4.3	172637	10	US-09-805-458A-3	Sequence 3, Appli
	30	17	4.3	326014	10	US-09-731-231A-3	Sequence 3, Appli
	31	17	4.3	1503841	9	US-09-946-807-1	Sequence 1, Appli
	32	17	4.3	1503841	10	US-09-795-668-1	Sequence 1, Appli
	33	17	4.3	1503841	10	US-09-795-686-1	Sequence 1, Appli
c	34	16	4.0	121	9	US-09-818-875-1321	Sequence 1321, Ap
	35	16	4.0	121	9	US-09-818-875-1322	Sequence 1322, Ap
c	36	16	4.0	121	9	US-09-818-875-1325	Sequence 1325, Ap
	37	16	4.0	121	9	US-09-818-875-1326	Sequence 1326, Ap
c	38	16	4.0	121	9	US-09-818-875-1329	Sequence 1329, Ap
	39	16	4.0	121	9	US-09-818-875-1330	Sequence 1330, Ap
c	40	16	4.0	121	9	US-09-818-875-1333	Sequence 1333, Ap
	41	16	4.0	121	9	US-09-818-875-1334	Sequence 1334, Ap
c	42	16	4.0	238	10	US-09-878-574-5259	Sequence 5259, Ap
c	43	16	4.0	255	10	US-09-878-574-13480	Sequence 13480, A
	44	16	4.0	281	10	US-09-964-824A-385	Sequence 385, App
	45	16	4.0	321	10	US-09-864-761-25530	Sequence 25530, A

ALIGNMENTS

```
RESULT 1
US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
;  GENERAL INFORMATION:
;    APPLICANT:  Hood, Leroy E.
;    APPLICANT:  Rowen, Lee
;    APPLICANT:  Koop, Ben F.
;    TITLE OF INVENTION:  DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U
;    NUMBER OF SEQUENCES:  1279
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/263,959
;      FILING DATE:  05-MAR-1999
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  920010.426C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  684973 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-263-959-1

  Query Match                 8.3%;  Score 33;  DB 10;  Length 684973;
```

```
  Best Local Similarity   100.0%;  Pred. No. 6.4e-08;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      66 TGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 98
           ||||||||||||||||||||||||||||||||
Db 404690 TGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAG 404658


RESULT 2
US-09-918-995-8549/c
; Sequence 8549, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8549
;   LENGTH: 361
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-8549

  Query Match             5.6%;  Score 22;  DB 9;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 0.063;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     149 GGCCCTGGTGGCATGGGCTCAC 170
           ||||||||||||||||||||||
Db     143 GGCCCTGGTGGCATGGGCTCAC 122


RESULT 3
US-09-974-879-27/c
; Sequence 27, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 125 Human Secreted Proteins
;  FILE REFERENCE: PZ020P2
;  CURRENT APPLICATION NUMBER: US/09/974,879
   CURRENT FILING DATE:  2001-10-12
   PRIOR APPLICATION NUMBER: US 60/239,893
   PRIOR FILING DATE: 2000-10-13
;  PRIOR APPLICATION NUMBER: US 09/818,683
;  PRIOR FILING DATE: 2001-03-28
;  PRIOR APPLICATION NUMBER: US 09/305,736
;  PRIOR FILING DATE: 1999-05-05
;  PRIOR APPLICATION NUMBER: PCT/US98/23435
;  PRIOR FILING DATE: 1998-11-04
;  PRIOR APPLICATION NUMBER: US 60/064,911
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,912
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,983
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,900
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,988
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,987
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,908
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,984
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/064,985
;  PRIOR FILING DATE: 1997-11-07
;  PRIOR APPLICATION NUMBER: US 60/066,094
;  PRIOR FILING DATE: 1997-11-17
;  PRIOR APPLICATION NUMBER: US 60/066,100
;  PRIOR FILING DATE: 1997-11-17
;  PRIOR APPLICATION NUMBER: US 60/066,089
;  PRIOR FILING DATE: 1997-11-17
;  PRIOR APPLICATION NUMBER: US 60/066,095
;  PRIOR FILING DATE: 1997-11-17
;  PRIOR APPLICATION NUMBER: US 60/066,090
;  PRIOR FILING DATE: 1997-11-17
;  NUMBER OF SEQ ID NOS: 611
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 802
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (337)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (359)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-974-879-27

  Query Match             4.8%;  Score 19;  DB 9;  Length 802;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     217 GTGGTTTCCTGGATGGGGT 235
           |||||||||||||||||||
Db     304 GTGGTTTCCTGGATGGGGT 286


RESULT 4
US-09-876-889-254/c
; Sequence 254, Application US/09876889
; Patent No. US20020076715A1
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION:  CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C3
;  CURRENT APPLICATION NUMBER: US/09/876,889
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 353
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 254
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(430)
;   OTHER INFORMATION: n = A,T,C or G
US-09-876-889-254

  Query Match             4.5%;  Score 18;  DB 10;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 GGGTTGCAAAGATCTGTG 208
           ||||||||||||||||||
Db     273 GGGTTGCAAAGATCTGTG 256


RESULT 5
US-09-866-562-80
; Sequence 80, Application US/09866562
```

```
; Patent No. US20020009758A1
; GENERAL INFORMATION:
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Klee, Jennifer
;  APPLICANT:  Switzer, Anne
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER.
;  FILE REFERENCE: 210121.502
;  CURRENT APPLICATION NUMBER: US/09/866,562
;  CURRENT FILING DATE:  2001-05-25
;  NUMBER OF SEQ ID NOS: 96
; SEQ ID NO 80
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 361,431
;   OTHER INFORMATION: n = A,T,C or G
US-09-866-562-80

  Query Match             4.5%;  Score 18;  DB 10;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 6
US-09-866-562-87
; Sequence 87, Application US/09866562
; Patent No. US20020009758A1
; GENERAL INFORMATION:
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Klee, Jennifer
;  APPLICANT:  Switzer, Anne
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER.
;  FILE REFERENCE: 210121.502
;  CURRENT APPLICATION NUMBER: US/09/866,562
;  CURRENT FILING DATE:  2001-05-25
;  NUMBER OF SEQ ID NOS: 96
;  SEQ ID NO 87
;  LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 361,431
;   OTHER INFORMATION: n = A,T,C or G
US-09-866-562-87

  Query Match             4.5%;  Score 18;  DB 10;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      91 TGCCTGAGTGGCTGCTCT 108
           ||||||||||||||||||
Db      18 TGCCTGAGTGGCTGCTCT 35


RESULT 7
US-09-864-761-9118/c
; Sequence 9118, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9118
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000053.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
US-09-864-761-9118

  Query Match             4.5%;  Score 18;  DB 10;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TGTGTGTGCCTGAGTGGC 102
           ||||||||||||||||||
Db     505 TGTGTGTGCCTGAGTGGC 488


RESULT 8
US-09-918-995-28049/c
```

; Sequence 28049, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28049
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(489)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28049

Query Match 4.3%; Score 17; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 325 TTTTTCTTCATTCTCTG 341
||||||||||||||||
Db 416 TTTTTCTTCATTCTCTG 400

RESULT 9
US-09-918-995-34650/c
; Sequence 34650, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34650
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(491)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-34650

Query Match 4.3%; Score 17; DB 9; Length 491;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 TTCTTCATTCTCTGTGG 344
||||||||||||||||
Db 138 TTCTTCATTCTCTGTGG 122

RESULT 10
US-09-918-995-19262/c
; Sequence 19262, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19262
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(492)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-19262

Query Match 4.3%; Score 17; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 CTCACTGCTTCCCTTGG 262
||||||||||||||||
Db 85 CTCACTGCTTCCCTTGG 69

RESULT 11
US-10-092-154-1899/c
; Sequence 1899, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1899
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1899

Query Match 4.3%; Score 17; DB 9; Length 531;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 TCTTTTTCATCTTTGCA 44
||||||||||||||||
Db 141 TCTTTTTCATCTTTGCA 125

RESULT 12
US-09-764-847-1899/c
; Sequence 1899, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1899
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1899

Query Match 4.3%; Score 17; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      28 TCTTTTTCATCTTTGCA 44
           |||||||||||||||||
Db     141 TCTTTTTCATCTTTGCA 125
```

RESULT 13
US-09-864-761-13668/c
; Sequence 13668, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13668
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z82201.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
US-09-864-761-13668

Query Match 4.3%; Score 17; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     327 TTTCTTCATTCTCTGTG 343
           |||||||||||||||||
Db     378 TTTCTTCATTCTCTGTG 362
```

RESULT 14
US-09-974-300-8084/c
; Sequence 8084, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8084
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-8084

Query Match 4.3%; Score 17; DB 10; Length 609;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      28 TCTTTTTCATCTTTGCA 44
           |||||||||||||||||
Db     536 TCTTTTTCATCTTTGCA 520
```

RESULT 15
US-09-774-639-99/c
; Sequence 99, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: PZ013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (486)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (934)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-774-639-99

Query Match 4.3%; Score 17; DB 9; Length 944;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     246 CTCACTGCTTCCCTTGG 262
           |||||||||||||||||
Db     515 CTCACTGCTTCCCTTGG 499
```

Search completed: April 25, 2003, 02:12:39
Job time : 286.231 secs

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2003, 22:30:38 ; Search time 1361 Seconds
(without alignments)
791.186 Million cell updates/sec

Title: US-09-513-999C-7869_COPY_1_37
Perfect score: 193
Sequence: 1 MGGSFALQDSFSSLQGLLGPEYVKLLGLCVCLSGCST 37

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09513999/runat_18042003_170936_28357/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999_@CGN_1_1_1687_@runat_18042003_170936_28357 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	193	100.0	169620	2	AC012674	AC012674 Homo sapi
c 2	193	100.0	199289	9	AC012378	AC012378 Homo sapi
c 3	129	66.8	152313	9	AL591594	AL591594 Human DNA
4	129	66.8	163314	2	AL590557	AL590557 Homo sapi
c 5	127	65.8	143372	9	AL137847	AL137847 Human DNA
c 6	117	60.6	148290	9	AC107979	AC107979 Homo sapi
c 7	117	60.6	165649	9	AC103996	AC103996 Homo sapi
c 8	117	60.6	169861	2	AC126324	AC126324 Homo sapi
c 9	117	60.6	192826	9	AC090762	AC090762 Homo sapi
10	116	60.1	131347	2	AC002421	AC002421 Homo sapi
11	116	60.1	170623	9	AL391375	AL391375 Human DNA
c 12	114	59.1	68314	2	AC126345	AC126345 Homo sapi
13	114	59.1	143717	2	AC069245	AC069245 Homo sapi
14	114	59.1	152354	2	AC016472	AC016472 Homo sapi
15	114	59.1	160066	2	AC027295	AC027295 Homo sapi
16	114	59.1	163475	9	AC078815	AC078815 Homo sapi
c 17	114	59.1	200543	2	AC016726	AC016726 Homo sapi
18	113	58.5	57662	2	AC107969	AC107969 Homo sapi
c 19	113	58.5	98360	9	HSDJ247C2	AL049713 Human DNA
c 20	113	58.5	145264	9	AC107939	AC107939 Homo sapi
21	112	58.0	147820	9	AC092837	AC092837 Homo sapi
c 22	112	58.0	199038	2	AC116565	AC116565 Homo sapi
c 23	109	56.5	86314	9	AC109592	AC109592 Homo sapi
c 24	109	56.5	159475	2	AC021378	AC021378 Homo sapi
c 25	109	56.5	176689	9	AL162414	AL162414 Human DNA
c 26	108	56.0	134760	9	AC099484	AC099484 Homo sapi
27	108	56.0	166706	9	AC068875	AC068875 Homo sapi
28	108	56.0	207408	2	AC068618	AC068618 Homo sapi
c 29	108	56.0	207548	9	AC087283	AC087283 Homo sapi
c 30	107	55.4	103138	2	AC115094	AC115094 Homo sapi
31	107	55.4	143200	9	AC008413	AC008413 Homo sapi
c 32	107	55.4	231948	2	AC113346	AC113346 Homo sapi
c 33	106.5	55.2	146059	2	AC019030	AC019030 Homo sapi
34	106	54.9	177447	2	AC104687	AC104687 Homo sapi
c 35	106	54.9	236281	9	AC004673	AC004673 Homo sapi
c 36	105	54.4	80417	9	AL512659	AL512659 Human DNA
37	105	54.4	108248	9	HS326I13	AL022158 Homo sapi
c 38	105	54.4	162778	9	AC009559	AC009559 Homo sapi
c 39	104	53.9	24725	9	AL392090	AL392090 Human DNA
40	104	53.9	120418	2	AP002509	AP002509 Homo sapi
41	104	53.9	147288	9	AL355812	AL355812 Human DNA
c 42	104	53.9	163231	9	AL162575	AL162575 Human DNA
c 43	104	53.9	175832	9	AC023080	AC023080 Homo sapi
44	104	53.9	187847	2	AC023155	AC023155 Homo sapi
45	102	52.8	66489	2	AC100766	AC100766 Homo sapi

ALIGNMENTS

RESULT 1

```
AC012674/c
LOCUS       AC012674              169620 bp    DNA     linear   HTG 07-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP1-458H3, WORKING DRAFT SEQUENCE,
            18 unordered pieces.
ACCESSION   AC012674
VERSION     AC012674.10  GI:9719580
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169620)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169620)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 7, 2000 this sequence version replaced gi:8705345.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMQG
                Center clone name: RP1-458H3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 139025 bases at least Q40
                Consensus quality: 154842 bases at least Q30
                Consensus quality: 159725 bases at least Q20
                Estimated insert size: 162720; sum-of-contigs estimation
                Estimated insert size: 171608; agarose-fp estimation
                Quality coverage: 3.9x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28689: contig of 28689 bp in length
            *    28690    28789: gap of unknown length
            *    28790    50832: contig of 22043 bp in length
            *    50833    50932: gap of unknown length
            *    50933    69144: contig of 18212 bp in length
            *    69145    69244: gap of unknown length
            *    69245    84204: contig of 14960 bp in length
            *    84205    84304: gap of unknown length
            *    84305    94667: contig of 10363 bp in length
            *    94668    94767: gap of unknown length
            *    94768   107261: contig of 12494 bp in length
            *   107262   107361: gap of unknown length
            *   107362   117550: contig of 10189 bp in length
            *   117551   117650: gap of unknown length
            *   117651   126939: contig of 9289 bp in length
            *   126940   127039: gap of unknown length
            *   127040   135040: contig of 8001 bp in length
            *   135041   135140: gap of unknown length
            *   135141   141639: contig of 6499 bp in length
            *   141640   141739: gap of unknown length
            *   141740   149558: contig of 7819 bp in length
            *   149559   149658: gap of unknown length
            *   149659   154562: contig of 4904 bp in length
            *   154563   154662: gap of unknown length
            *   154663   158987: contig of 4325 bp in length
            *   158988   159087: gap of unknown length
            *   159088   162376: contig of 3289 bp in length
            *   162377   162476: gap of unknown length
            *   162477   165191: contig of 2715 bp in length
            *   165192   165291: gap of unknown length
            *   165292   167173: contig of 1882 bp in length
            *   167174   167273: gap of unknown length
            *   167274   168393: contig of 1120 bp in length
            *   168394   168493: gap of unknown length
            *   168494   169620: contig of 1127 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169620
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP1-458H3"
BASE COUNT    52024 a  33180 c  32128 g  50322 t   1966 others
ORIGIN

Alignment Scores:
Pred. No.:              9.18e-18        Length:         169620
Score:                  193.00          Matches:        37
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC012674 (1-169620)

Qy      1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87441 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 87382

Qy     21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87381 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 87331

RESULT 2
AC012378/c
LOCUS       AC012378              199289 bp    DNA     linear   PRI 09-AUG-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-420M1 map 15q21.3, complete
            sequence.
ACCESSION   AC012378
VERSION     AC012378.10  GI:15145648
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199289)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
```

```
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199289)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Madan,A.,
            Owen,M.P., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 199289)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:13775292.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399


            Note: Data from overlapping clones AC022083 [Drafting center:
            UWMSC], AC012674 [Drafting center: BCM] and AC009997 [drafting
            center: UWMSC] were added for finishing.
FEATURES             Location/Qualifiers
     source          1. .199289
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q21.3"
                     /clone="RP11-420M1"
                     /clone_lib="RPCI human bac library 11"
                     /note="This clone overlaps CTD-2137J4 AC022083, RP11-458H3
                     AC012674 and RP11-291H24 AC009997. Data from overlapping
                     BACs were added and the consensus sequence determined from
                     RP11-420M1 to the extent possible."
     misc_feature    1. .165029
                     /note="overlap with RP11-458H3, AC012674"
     misc_feature    1. .11647
                     /note="overlap with CTD-2137J4, AC022083"
     unsure          49933. .50105
                     /note="low quality data"
     unsure          52173. .52233
                     /note="low quality data"
     unsure          52381. .52385
                     /note="low quality data"
     unsure          52412. .52483
                     /note="low quality data"
     unsure          52541. .52570
                     /note="low quality data"
     unsure          53242. .53257
                     /note="low quality data"
     misc_feature    53940. .54690
                     /note="sequence data generated from subcloned PCR product"
     unsure          72103. .72105
                     /note="low quality data"
     unsure          155032. .155038
                     /note="low quality data"
     misc_feature    166787. .199289
                     /note="overlap with RP11-291H24, AC009997"
BASE COUNT    63165 a  40338 c  36951 g  58835 t
ORIGIN

Alignment Scores:

Pred. No.:                1.08e-17      Length:          199289
Score:                    193.00        Matches:         37
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC012378 (1-199289)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14756 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 14697

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14696 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 14646

RESULT 3
AL591594/c
LOCUS       AL591594              152313 bp    DNA     linear   PRI 16-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-424N15 on chromosome 1, complete
            sequence.
ACCESSION   AL591594
VERSION     AL591594.9  GI:16944148
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152313)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16605741.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-424N15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-424N15 The true
            right end of clone RP11-518L10 is at 64264 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .152313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-424N15"
                     /clone_lib="RPCI-11.2"
     misc_feature    18821. .18877
                     /note="Sequence from overlapping clone RP11-24C8
```

```
                (AL590557). Assembly confirmed by restriction digest."
BASE COUNT    45419 a  29865 c  29040 g  47989 t
ORIGIN

Alignment Scores:
Pred. No.:              5.03e-08        Length:         152313
Score:                  129.00          Matches:        23
Percent Similarity:     92.86%          Conservative:   3
Best Local Similarity:  82.14%          Mismatches:     2
Query Match:            66.84%          Indels:         0
DB:                     9               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AL591594 (1-152313)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||   |||||:::|||   |||:::||||||||||:::|||||||||||||||||||||
Db 100967 GACAAATTTTCCACCTTGGCAGGGGTCCTGGGTCCACAGTATGTAAAACTCCTGGGCCTC 100908

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||||||||||||||||||||||
Db 100907 TGCGTGTGCCTGAGTGGCTGCTCT 100884

RESULT 4
AL590557
LOCUS       AL590557              163314 bp    DNA     linear   HTG 20-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-24C8, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL590557
VERSION     AL590557.8  GI:13992136
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163314)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 8, 2001 this sequence version replaced gi:13990622.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA24C8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 161385 bases at least Q40
            Consensus quality: 161940 bases at least Q30
            Consensus quality: 162264 bases at least Q20
            Insert size: 162614; sum-of-contigs
            Insert size: 164357; 4.9% error; agarose-fp
            Quality coverage: 6.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.67x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    16091: contig of 16091 bp in length
            *   16092 16191: gap of      100 bp
            *   16192    51184: contig of 34993 bp in length
            *   51185 51284: gap of      100 bp
            *   51285    67941: contig of 16657 bp in length
            *   67942 68041: gap of      100 bp
            *   68042   103921: contig of 35880 bp in length
            *  103922 104021: gap of      100 bp
            *  104022   107020: contig of 2999 bp in length
            *  107021 107120: gap of      100 bp
            *  107121   123748: contig of 16628 bp in length
            *  123749 123848: gap of      100 bp
            *  123849   137091: contig of 13243 bp in length
            *  137092 137191: gap of      100 bp
            *  137192   163314: contig of 26123 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163314
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-24C8"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .16091
                     /note="assembly_fragment:00603
                     fragment_chain:1"
     misc_feature    16192. .51184
                     /note="assembly_fragment:01328
                     fragment_chain:1"
     misc_feature    51285. .67941
                     /note="assembly_fragment:01579
                     fragment_chain:1"
     misc_feature    68042. .103921
                     /note="assembly_fragment:00704
                     fragment_chain:1"
     misc_feature    104022. .107020
                     /note="assembly_fragment:02225
                     fragment_chain:1"
     misc_feature    107121. .123748
                     /note="assembly_fragment:02172
                     fragment_chain:1"
     misc_feature    123849. .137091
                     /note="assembly_fragment:01998
                     fragment_chain:1"
     misc_feature    137192. .163314
                     /note="assembly_fragment:03121"
BASE COUNT    51215 a  31325 c  32024 g  48050 t    700 others
ORIGIN

Alignment Scores:
Pred. No.:              5.39e-08        Length:         163314
Score:                  129.00          Matches:        23
Percent Similarity:     92.86%          Conservative:   3
Best Local Similarity:  82.14%          Mismatches:     2
Query Match:            66.84%          Indels:         0
DB:                     2               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AL590557 (1-163314)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||   |||||:::|||   |||:::||||||||||:::|||||||||||||||||||||
Db  55961 GACAAATTTTCCACCTTGGCAGGGGTCCTGGGTCCACAGTATGTAAAACTCCTGGGCCTC 56020

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||||||||||||||||||||||
Db  56021 TGCGTGTGCCTGAGTGGCTGCTCT 56044

RESULT 5
AL137847/c
LOCUS       AL137847              143372 bp    DNA     linear   PRI 16-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-439K3 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL137847
VERSION     AL137847.12  GI:16973786
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 143372)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 17, 2001 this sequence version replaced gi:16408610.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-439K3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-439K3 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-439K3 is at 1 in this sequence. The
            true left end of clone RP11-344L17 is at 141373 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .143372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-439K3"
                     /clone_lib="RPCI-11.2"
BASE COUNT    42809 a  30224 c  28152 g  42187 t
ORIGIN

Alignment Scores:
Pred. No.:              9.58e-08        Length:         143372
Score:                  127.00          Matches:        23
Percent Similarity:     89.29%          Conservative:   2
Best Local Similarity:  82.14%          Mismatches:     3
Query Match:            65.80%          Indels:         0
DB:                     9               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AL137847 (1-143372)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||   :::|||   |||:::||||||||||||||||||||||||||||||||
Db   82342 GATAGATTTCCTGCCTTGCCAGGGATCCTGGGGCCAGAGTATGTAAAACTCCTGGGTCTC 82283

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||||||||||||||||||||||
Db   82282 TGTGTGTGCCTGAGTGGCTGCTCT 82259

RESULT 6
AC107979/c
LOCUS       AC107979              148290 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone CTD-3049M7, complete sequence.
ACCESSION   AC107979
VERSION     AC107979.7  GI:21592043
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-3049M7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 148290)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 148290)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
```

```
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 26, 2002 this sequence version replaced gi:21321840.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24533
                Center clone name: 3049_M_7
            --------------.
FEATURES             Location/Qualifiers
     source          1. .148290
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-3049M7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   complement(1982. .2092)
                     /rpt_family="MIR"
     repeat_region   complement(3356. .3654)
                     /rpt_family="AluSx"
     repeat_region   3789. .3844
                     /rpt_family="MER58"
     repeat_region   complement(4429. .4728)
                     /rpt_family="AluSp"
     repeat_region   5494. .5723
                     /rpt_family="(TA)n"
     repeat_region   5727. .5766
                     /rpt_family="(GA)n"
     repeat_region   5806. .5857
                     /rpt_family="AT_rich"
     repeat_region   6377. .6581
                     /rpt_family="L1"
     repeat_region   6647. .7069
                     /rpt_family="L1ME3"
     repeat_region   7057. .7631
                     /rpt_family="L1ME3"
     repeat_region   complement(8821. .8958)
                     /rpt_family="MIR3"
     repeat_region   11129. .11224
                     /rpt_family="FRAM"
     repeat_region   11229. .11338
                     /rpt_family="CT-rich"
     repeat_region   11943. .12090
                     /rpt_family="MIR"
     repeat_region   12818. .13060
                     /rpt_family="MIR"
     repeat_region   13118. .19146
                     /rpt_family="L1PA2"
     repeat_region   20846. .20912
                     /rpt_family="CT-rich"
     repeat_region   21294. .21314
                     /rpt_family="AT_rich"
     repeat_region   21635. .23417
                     /rpt_family="L1PA16"
     repeat_region   23438. .25940
                     /rpt_family="L1PA16"
     repeat_region   26121. .26145
                     /rpt_family="(TTCA)n"
     repeat_region   26293. .26417
                     /rpt_family="MIR"
     repeat_region   complement(26872. .26937)
                     /rpt_family="L2"
     repeat_region   complement(27064. .27431)
                     /rpt_family="L2"
     repeat_region   27917. .28357
                     /rpt_family="MLT1B"
     repeat_region   28460. .28806
                     /rpt_family="MLT1I"
     repeat_region   29509. .29698
                     /rpt_family="MIR"
     repeat_region   29716. .29790
                     /rpt_family="CT-rich"
     repeat_region   29791. .29811
                     /rpt_family="AT_rich"
     repeat_region   complement(31182. .31787)
                     /rpt_family="L2"
     repeat_region   complement(32398. .32581)
                     /rpt_family="AluSq"
     repeat_region   complement(32582. .32777)
                     /rpt_family="L1PB1"
     repeat_region   complement(32778. .32912)
                     /rpt_family="AluSq"
     repeat_region   33104. .33207
                     /rpt_family="L1MC5"
     repeat_region   complement(33208. .33365)
                     /rpt_family="MER53"
     repeat_region   33366. .33529
                     /rpt_family="L1MC5"
     repeat_region   33877. .34506
                     /rpt_family="L1MEd"
     repeat_region   34667. .34808
                     /rpt_family="L1MEd"
     repeat_region   35585. .36698
                     /rpt_family="L1M4"
     repeat_region   complement(36699. .36830)
                     /rpt_family="FLAM_C"
     repeat_region   36831. .38141
                     /rpt_family="L1M4"
     repeat_region   38194. .38373
                     /rpt_family="MLT1I"
     repeat_region   39530. .39757
                     /rpt_family="MIR"
     repeat_region   39992. .40017
                     /rpt_family="(TTTTG)n"
     repeat_region   40283. .40583
                     /rpt_family="AluY"
     repeat_region   41414. .41616
                     /rpt_family="MLT1C"

Alignment Scores:
Pred. No.:              3.35e-06        Length:         148290
Score:                  117.00          Matches:        21
Percent Similarity:     85.71%          Conservative:   3
Best Local Similarity:  75.00%          Mismatches:     4
Query Match:            60.62%          Indels:         0
DB:                     9               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC107979 (1-148290)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||       |||:::       |||:::|||||||||||||||:::|||||||||||||||
Db  84542 GATGGATCTTCCACCTCACTGGGGATCCTGGGGCCAGAGTATATAAAACTCCTGGGTCTC 84483

Qy     29 CysValCysLeuSerGlyCysSer 36
          ||||||||||||||||||||||||
Db  84482 TGTGTGTGCCTGAGTGGCTGCTCA 84459

RESULT 7
AC103996/c
LOCUS       AC103996               165649 bp    DNA     linear   PRI 01-JUL-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-76E17, complete sequence.
```

```
ACCESSION   AC103996
VERSION     AC103996.7  GI:21637504
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-76E17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165649)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 165649)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 1, 2002 this sequence version replaced gi:21592191.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21917
                Center clone name: 76_E_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .165649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-76E17"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(2. .865)
                     /rpt_family="L1PA13"
     repeat_region   complement(864. .894)
                     /rpt_family="L1PA13"
     repeat_region   complement(895. .1248)
                     /rpt_family="THE1A"
     repeat_region   complement(1249. .1504)
                     /rpt_family="L1PA13"
     repeat_region   1505. .1626
                     /rpt_family="AluSx"
     repeat_region   1627. .1659
                     /rpt_family="(CAAA)n"
     repeat_region   1660. .1831
                     /rpt_family="AluSx"
     repeat_region   complement(1832. .3132)
                     /rpt_family="L1PA13"
     repeat_region   complement(3140. .3699)
                     /rpt_family="L1M4c"
     repeat_region   complement(4461. .4755)
                     /rpt_family="AluSx"
     repeat_region   5027. .5397
                     /rpt_family="L2"
     repeat_region   complement(5650. .6330)
                     /rpt_family="L1ME1"
     repeat_region   complement(6343. .6460)
                     /rpt_family="L1ME1"
     repeat_region   7214. .7319
                     /rpt_family="L1MC3"
     repeat_region   7354. .7655
                     /rpt_family="L1MC3"
     repeat_region   7659. .7743
                     /rpt_family="L1PA10"
     repeat_region   complement(7745. .7843)
                     /rpt_family="AluSp/q"
     repeat_region   7844. .14156
                     /rpt_family="L1PA10"
     repeat_region   14198. .14328
                     /rpt_family="AluY"
     repeat_region   14340. .14503
                     /rpt_family="(TA)n"
     repeat_region   14507. .15032
                     /rpt_family="L1MC3"
     repeat_region   15036. .15220
                     /rpt_family="L1MC3"
```

```
     repeat_region   complement(15218. .16449)
                     /rpt_family="L1PA4"
     repeat_region   16450. .18154
                     /rpt_family="L1PA4"
     repeat_region   18155. .19210
                     /rpt_family="L1MC3"
     repeat_region   19211. .19266
                     /rpt_family="(TA)n"
     repeat_region   19267. .19372
                     /rpt_family="L1MC3"
     repeat_region   19401. .19476
                     /rpt_family="(TTATA)n"
     repeat_region   19483. .19546
                     /rpt_family="(CATATA)n"
     repeat_region   19596. .19625
                     /rpt_family="AT_rich"
     repeat_region   19696. .19750
                     /rpt_family="GA-rich"
     repeat_region   19752. .19920
                     /rpt_family="L1MD3"
     repeat_region   19994. .20102
                     /rpt_family="L2"
     unsure          20049. .20114
                     /note="single clone coverage"
     repeat_region   20485. .20655
                     /rpt_family="MIR3"
     repeat_region   complement(21285. .21441)
                     /rpt_family="MIR"
     repeat_region   21496. .21717
                     /rpt_family="L2"
     repeat_region   complement(21720. .21796)
                     /rpt_family="MIR"
     repeat_region   complement(21943. .22131)
                     /rpt_family="MIR"
     repeat_region   23082. .23195
                     /rpt_family="L2"
     repeat_region   23198. .23248
                     /rpt_family="GA-rich"
     repeat_region   23267. .23333
                     /rpt_family="(CATA)n"
     repeat_region   complement(24465. .24833)
                     /rpt_family="MLT1A2"
     repeat_region   26142. .26334
                     /rpt_family="MIR"
     repeat_region   complement(27659. .27811)
                     /rpt_family="MIR"
     repeat_region   complement(28811. .28880)
                     /rpt_family="MIR"
     repeat_region   complement(29026. .29269)
                     /rpt_family="MIR"
     repeat_region   complement(29673. .29706)
                     /rpt_family="MSTB"

Alignment Scores:
Pred. No.:              3.74e-06        Length:         165649
Score:                  117.00          Matches:        21
Percent Similarity:     85.71%          Conservative:   3
Best Local Similarity:  75.00%          Mismatches:     4
Query Match:            60.62%          Indels:         0
DB:                     9               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC103996 (1-165649)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||      |||:::      |||:::|||||||||||||||:::|||||||||||||||
Db   8153 GATGGATCTTCCACCTCACTGGGGATCCTGGGGCCAGAGTATATAAAACTCCTGGGTCTC 8094

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||||||||||||||||||||||||
Db   8093 TGTGTGTGCCTGAGTGGCTGCTCA 8070

RESULT 8
AC126324/c
```

```
LOCUS       AC126324              169861 bp    DNA     linear   HTG 06-AUG-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-358H20 map 11, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC126324
VERSION     AC126324.2  GI:22123018
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169861)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-358H20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169861)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 169861)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 6, 2002 this sequence version replaced gi:21699255.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L27749
                Center clone name: 358_H_20
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
```

```
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    71322: contig of 71322 bp in length
            *   71323 71422: gap of      100 bp
            *   71423   169861: contig of 98439 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169861
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-358H20"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    55771 a  33379 c  31394 g  49217 t    100 others
ORIGIN
[redacted]gnment Scores:
Pred. No.:              3.83e-06       Length:         169861
Score:                  117.00         Matches:        23
Percent Similarity:     79.41%         Conservative:   4
Best Local Similarity:  67.65%         Mismatches:     7
Query Match:            60.62%         Indels:         0
DB:                     2              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC126324 (1-169861)

Qy      3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
          |||      ||||||   :::      :::||||||   |||:::|||||||||||||:::
Db  73373 GGACTTCACGCCCTGCCTGAGTTGCAGACAAGTTTGTTGGGGATCCTTGGGCCAGAGTTT 73314

Qy     23 ValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
          |||||||||||||||||||||||||||||||   ||||||
Db  73313 GTAAAGCTCCTGGGTCTCTGTGTGTGTCTGAGCAGCTGCTCT 73272

RESULT 9
AC090762/c
LOCUS       AC090762               192826 bp    DNA     linear   PRI 28-FEB-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-387E8, complete sequence.
ACCESSION   AC090762
VERSION     AC090762.9  GI:18997378
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[redacted]ERENCE   1  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-387E8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 192826)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2002 this sequence version replaced gi:18377189.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12392
                Center clone name: 387_E_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192826
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-387E8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(864. .1222)
                     /rpt_family="L1PB3"
     repeat_region   1223. .1252
                     /rpt_family="(TAGG)n"
     repeat_region   complement(1253. .1492)
                     /rpt_family="L1PB3"
     repeat_region   4287. .4576
                     /rpt_family="AluJb"
     repeat_region   4671. .5034
                     /rpt_family="THE1B"
     repeat_region   5130. .5438
                     /rpt_family="AluSx"
     repeat_region   5517. .5662
                     /rpt_family="MIR"
     repeat_region   5798. .5827
                     /rpt_family="AT_rich"
     repeat_region   5985. .6114
                     /rpt_family="AluJb"
     repeat_region   6227. .6253
                     /rpt_family="(CA)n"
```

```
repeat_region   7329. .7446
                /rpt_family="MIR"
repeat_region   complement(8147. .8452)
                /rpt_family="AluJo"
repeat_region   8540. .8639
                /rpt_family="MER45"
repeat_region   complement(8832. .8916)
                /rpt_family="L2"
repeat_region   10391. .10599
                /rpt_family="MER3"
repeat_region   11688. .11828
                /rpt_family="L1MC/D"
repeat_region   12029. .12078
                /rpt_family="AT_rich"
repeat_region   12092. .12447
                /rpt_family="THE1C"
repeat_region   complement(13616. .13751)
                /rpt_family="MIR3"
repeat_region   complement(13958. .14142)
                /rpt_family="MIR"
repeat_region   complement(15277. .15553)
                /rpt_family="MER8"
repeat_region   complement(15811. .16005)
                /rpt_family="MIR"
repeat_region   complement(16101. .16440)
                /rpt_family="L3"
repeat_region   16920. .16958
                /rpt_family="(TCCC)n"
repeat_region   complement(17145. .17444)
                /rpt_family="AluSx"
repeat_region   complement(18418. .19953)
                /rpt_family="L1MEc"
repeat_region   complement(19978. .20262)
                /rpt_family="L1MEc"
repeat_region   complement(20288. .20794)
                /rpt_family="L1MEc"
repeat_region   complement(20822. .21097)
                /rpt_family="L1MEc"
repeat_region   complement(21345. .21743)
                /rpt_family="L1MEc"
repeat_region   22599. .23518
                /rpt_family="L1MEc"
repeat_region   23527. .23901
                /rpt_family="L1MEc"
repeat_region   complement(23927. .24026)
                /rpt_family="MSTB1"
repeat_region   complement(24027. .24256)
                /rpt_family="MER30"
repeat_region   complement(24257. .24557)
                /rpt_family="MSTB1"
repeat_region   24563. .24594
                /rpt_family="AT_rich"
repeat_region   complement(25258. .25639)
                /rpt_family="L1MCc"
repeat_region   complement(25688. .25838)
                /rpt_family="L1MCc"
repeat_region   26167. .26506
                /rpt_family="THE1B"
repeat_region   26746. .27094
                /rpt_family="Tigger2a"
repeat_region   27095. .27184
                /rpt_family="MADE1"
repeat_region   27185. .27279
                /rpt_family="Tigger2a"
repeat_region   28321. .28363
                /rpt_family="(TG)n"
repeat_region   29172. .29333
                /rpt_family="MIR"
repeat_region   30664. .30898
                /rpt_family="L1MB8"
repeat_region   31577. .31598
                /rpt_family="AT_rich"
repeat_region   32378. .32475
                /rpt_family="CT-rich"
repeat_region   34655. .34960
                /rpt_family="AluY"
repeat_region   complement(35296. .35389)
                /rpt_family="MER5B"
repeat_region   complement(35419. .35517)
                /rpt_family="L1MC4"
repeat_region   complement(35518. .35823)
                /rpt_family="AluSx"
repeat_region   complement(35824. .36030)
                /rpt_family="L1MC4"
repeat_region   36682. .36758
                /rpt_family="MER5A"
repeat_region   36975. .37160
                /rpt_family="MER5A"
repeat_region   38140. .38353
                /rpt_family="MER96B"
repeat_region   38633. .38656
                /rpt_family="(TTTA)n"
repeat_region   complement(38657. .40056)
                /rpt_family="L1PA5"
repeat_region   complement(40060. .40250)

Alignment Scores:
Pred. No.:              4.34e-06        Length:         192826
Score:                  117.00          Matches:        21
Percent Similarity:     85.71%          Conservative:   3
Best Local Similarity:  75.00%          Mismatches:     4
Query Match:            60.62%          Indels:         0
DB:                     9               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC090762 (1-192826)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||      |||:::      |||:::||||||||||||||||:::||||||||||||||||
Db 190989 GATGGATCTTCCACCTCACTGGGGATCCTGGGGCCAGAGTATATAAAACTCCTGGGTCTC 190930

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||||||||||||||||||||||||
Db 190929 TGTGTGTGCCTGAGTGGCTGCTCA 190906

RESULT 10
AC002421
LOCUS       AC002421              131347 bp    DNA     linear   HTG 13-JUN-2002
DEFINITION  Homo sapiens chromosome X clone pWXD1, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC002421
VERSION     AC002421.2  GI:21405641
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131347)
  AUTHORS   Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and
            Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 131347)
  AUTHORS   Brownstein,B.H., States,D.J. and Mazzarella,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
            Washington University School of Medicine, 4566 Scott Avenue, St.
            Louis, MO 63110, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:2323248.
            Current status of this project is available at:
            'http://genome.wustl.edu/cgm/seq_projects.html'
            Submitted by:
            Ellson Chen,
            Advanced Center for Genetic Technology,
            Applied Biosystems Division of Perlin Elmer Corp.,
            850 Lincoln Center Drive,
            Foster City, CA 94404 USA
```

```
            e-mail: ellson@genseq.apldbio.com

            and

            Buddy Brownstein,
            Center for Genetics in Medicine,
            Washington University School of Medicine, Box 8232
            4566 Scott Avenue,
            St. Louis, MO 63110, USA
            e-mail: buddy@genetics.wustl.edu

            and

            David J. States,
            Institute for Biomedical Computing
            Washington University in St. Louis
            700 South Euclid Ave.
            St. Louis, MO 63108 USA
            e-mail: states@ibc.wustl.edu.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    22595: contig of 22595 bp in length
            *    22596 22695: gap of      100 bp
            *    22696    66983: contig of 44288 bp in length
            *    66984 67083: gap of      100 bp
            *    67084   122847: contig of 55764 bp in length
            *   122848 122947: gap of      100 bp
            *   122948   131347: contig of 8400 bp in length.
FEATURES             Location/Qualifiers
     source          1. .131347
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="pWXD1"
BASE COUNT    40590 a  23040 c  23923 g  43494 t    300 others
ORIGIN

Alignment Scores:
Pred. No.:              4.22e-06       Length:         131347
Score:                  116.00         Matches:        21
Percent Similarity:     86.21%         Conservative:   4
Best Local Similarity:  72.41%         Mismatches:     4
Query Match:            60.10%         Indels:         0
                        2              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC002421 (1-131347)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||      :::::|||   |||:::|||||||||||||||||||||||||||||||||
Db 103911 GATGGATCTGCCACCTTGCTGGGAATCCTGGGGCCAGAGTATGTAAAACTTCTGGGTCTC 103970

Qy     29 CysValCysLeuSerGlyCysSerThr 37
          |||||||||||||||   ||||||:::
Db 103971 TGTGTGTGCCTGAGCAGCTGCTCTTCT 103997

RESULT 11
AL391375
LOCUS       AL391375              170623 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-375A20 on chromosome X, complete
            sequence.
ACCESSION   AL391375
VERSION     AL391375.11  GI:12657182
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170623)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 2, 2001 this sequence version replaced gi:11414631.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-375A20 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-375A20.
FEATURES             Location/Qualifiers
     source          1. .170623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-375A20"
                     /clone_lib="RPCI-11.2"
     misc_feature    71147. .71285
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
BASE COUNT    53145 a  29784 c  30919 g  56775 t
ORIGIN

Alignment Scores:
Pred. No.:              5.47e-06       Length:         170623
Score:                  116.00         Matches:        21
Percent Similarity:     86.21%         Conservative:   4
Best Local Similarity:  72.41%         Mismatches:     4
Query Match:            60.10%         Indels:         0
DB:                     9              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AL391375 (1-170623)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||      :::::|||   |||:::|||||||||||||||||||||||||||||||||
Db  99926 GATGGATCTGCCACCTTGCTGGGAATCCTGGGGCCAGAGTATGTAAAACTTCTGGGTCTC 99985

Qy     29 CysValCysLeuSerGlyCysSerThr 37
          |||||||||||||||   ||||||:::
Db  99986 TGTGTGTGCCTGAGCAGCTGCTCTTCT 100012

RESULT 12
AC126345/c
LOCUS       AC126345              68314 bp    DNA     linear   HTG 30-JUL-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-100E23 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC126345
VERSION     AC126345.2  GI:22004319
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68314)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-100E23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68314)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 68314)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
   ITLE     Direct Submission
   OURNAL   Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 30, 2002 this sequence version replaced gi:21699290.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L27743
                Center clone name: 100_E_23
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      728: contig of 728 bp in length
            *      729 828: gap of      100 bp
            *      829     1544: contig of 716 bp in length
            *     1545 1644: gap of      100 bp
            *     1645     2360: contig of 716 bp in length
            *     2361 2460: gap of      100 bp
            *     2461     3152: contig of 692 bp in length
            *     3153 3252: gap of      100 bp
            *     3253     3960: contig of 708 bp in length
            *     3961 4060: gap of      100 bp
            *     4061     4774: contig of 714 bp in length
            *     4775 4874: gap of      100 bp
            *     4875     5583: contig of 709 bp in length
            *     5584 5683: gap of      100 bp
            *     5684     6389: contig of 706 bp in length
            *     6390 6489: gap of      100 bp
            *     6490     7213: contig of 724 bp in length
            *     7214 7313: gap of      100 bp
            *     7314     8019: contig of 706 bp in length
            *     8020 8119: gap of      100 bp
            *     8120     8851: contig of 732 bp in length
            *     8852 8951: gap of      100 bp
            *     8952     9622: contig of 671 bp in length
            *     9623 9722: gap of      100 bp
            *     9723    10434: contig of 712 bp in length
            *    10435 10534: gap of      100 bp
            *    10535    11257: contig of 723 bp in length
            *    11258 11357: gap of      100 bp
            *    11358    12070: contig of 713 bp in length
            *    12071 12170: gap of      100 bp
            *    12171    12895: contig of 725 bp in length
            *    12896 12995: gap of      100 bp
            *    12996    13714: contig of 719 bp in length
            *    13715 13814: gap of      100 bp
            *    13815    14524: contig of 710 bp in length
            *    14525 14624: gap of      100 bp
            *    14625    15335: contig of 711 bp in length
            *    15336 15435: gap of      100 bp
            *    15436    16165: contig of 730 bp in length
            *    16166 16265: gap of      100 bp
            *    16266    16986: contig of 721 bp in length
            *    16987 17086: gap of      100 bp
            *    17087    17817: contig of 731 bp in length
            *    17818 17917: gap of      100 bp
            *    17918    18633: contig of 716 bp in length
            *    18634 18733: gap of      100 bp
            *    18734    19459: contig of 726 bp in length
            *    19460 19559: gap of      100 bp
            *    19560    20266: contig of 707 bp in length
            *    20267 20366: gap of      100 bp
            *    20367    21090: contig of 724 bp in length
            *    21091 21190: gap of      100 bp
            *    21191    21901: contig of 711 bp in length
            *    21902 22001: gap of      100 bp
            *    22002    22721: contig of 720 bp in length
            *    22722 22821: gap of      100 bp
            *    22822    23530: contig of 709 bp in length
            *    23531 23630: gap of      100 bp
            *    23631    24336: contig of 706 bp in length
            *    24337 24436: gap of      100 bp
            *    24437    25167: contig of 731 bp in length
            *    25168 25267: gap of      100 bp
            *    25268    25989: contig of 722 bp in length
            *    25990 26089: gap of      100 bp
            *    26090    26816: contig of 727 bp in length
            *    26817 26916: gap of      100 bp
            *    26917    27618: contig of 702 bp in length
            *    27619 27718: gap of      100 bp
            *    27719    28439: contig of 721 bp in length
            *    28440 28539: gap of      100 bp
            *    28540    29257: contig of 718 bp in length
            *    29258 29357: gap of      100 bp
```

```
*    29358    30053: contig of 696 bp in length
*    30054 30153: gap of      100 bp
*    30154    30856: contig of 703 bp in length
*    30857 30956: gap of      100 bp
*    30957    31680: contig of 724 bp in length
*    31681 31780: gap of      100 bp
*    31781    32509: contig of 729 bp in length
*    32510 32609: gap of      100 bp
*    32610    33331: contig of 722 bp in length
*    33332 33431: gap of      100 bp
*    33432    34153: contig of 722 bp in length
*    34154 34253: gap of      100 bp
*    34254    34964: contig of 711 bp in length
*    34965 35064: gap of      100 bp
*    35065    35759: contig of 695 bp in length
*    35760 35859: gap of      100 bp
*    35860    36566: contig of 707 bp in length
*    36567 36666: gap of      100 bp
*    36667    37385: contig of 719 bp in length
*    37386 37485: gap of      100 bp
*    37486    38202: contig of 717 bp in length
*    38203 38302: gap of      100 bp
*    38303    39019: contig of 717 bp in length
*    39020 39119: gap of      100 bp
*    39120    39833: contig of 714 bp in length
*    39834 39933: gap of      100 bp
*    39934    40657: contig of 724 bp in length
*    40658 40757: gap of      100 bp
*    40758    41471: contig of 714 bp in length
*    41472 41571: gap of      100 bp
*    41572    42284: contig of 713 bp in length
*    42285 42384: gap of      100 bp
*    42385    43100: contig of 716 bp in length
*    43101 43200: gap of      100 bp
*    43201    43926: contig of 726 bp in length
*    43927 44026: gap of      100 bp
*    44027    44745: contig of 719 bp in length
*    44746 44845: gap of      100 bp
*    44846    45552: contig of 707 bp in length
*    45553 45652: gap of      100 bp
*    45653    46336: contig of 684 bp in length
*    46337 46436: gap of      100 bp
*    46437    47128: contig of 692 bp in length
*    47129 47228: gap of      100 bp
*    47229    47951: contig of 723 bp in length
*    47952 48051: gap of      100 bp

Alignment Scores:
Seq. No.:              4.46e-06        Length:         68314
Score:                 114.00          Matches:        22
Percent Similarity:    83.87%          Conservative:   4
Best Local Similarity: 70.97%          Mismatches:     5
Query Match:           59.07%          Indels:         0
DB:                    2               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC126345 (1-68314)

Qy      6 AlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeu 25
          ||||||   :::      :::||||||   |||:::||||||||||||:::|||||||||
Db  43053 GCCCTGCCTGAGTTGCAGACAAGTTTGTTGGGGATCCTTGGGCCCGAGTTTGTAAAGCTC 42994

Qy     26 LeuGlyLeuCysValCysLeuSerGlyCysSer 36
          ||||||||||||||||||||||||   ||||||
Db  42993 CTGGGTCTCTGTGTGTGTCTGAGCAGCTGCTCT 42961

RESULT 13
AC069245
LOCUS       AC069245               143717 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-757N13, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC069245
VERSION     AC069245.10  GI:20335692
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143717)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143717)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 143717)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:18449608.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBNX
                Center clone name: RP11-757N13
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 19% of reads
                Chemistry: Dye-terminator Big Dye: 81% of reads
                Assembly program: Phrap; version 0.990329
```

```
            Consensus quality: 133223 bases at least Q40
            Consensus quality: 141119 bases at least Q30
            Consensus quality: 145567 bases at least Q20
            Estimated insert size: 148183; sum-of-contigs estimation
            Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
        --------------
        * NOTE: Estimated insert size may differ from sequence length
        *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 17 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     2357: contig of 2357 bp in length
        *     2358     2457: gap of unknown length
        *     2458     4966: contig of 2509 bp in length
        *     4967     5066: gap of unknown length
        *     5067     7336: contig of 2270 bp in length
        *     7337     7436: gap of unknown length
        *     7437    10419: contig of 2983 bp in length
        *    10420    10519: gap of unknown length
        *    10520    14814: contig of 4295 bp in length
        *    14815    14914: gap of unknown length
        *    14915    18175: contig of 3261 bp in length
        *    18176    18275: gap of unknown length
        *    18276    21287: contig of 3012 bp in length
        *    21288    21387: gap of unknown length
        *    21388    25727: contig of 4340 bp in length
        *    25728    25827: gap of unknown length
        *    25828    29882: contig of 4055 bp in length
        *    29883    29982: gap of unknown length
        *    29983    35515: contig of 5533 bp in length
        *    35516    35615: gap of unknown length
        *    35616    40650: contig of 5035 bp in length
        *    40651    40750: gap of unknown length
        *    40751    47604: contig of 6854 bp in length
        *    47605    47704: gap of unknown length
        *    47705    59212: contig of 11508 bp in length
        *    59213    59312: gap of unknown length
        *    59313    74599: contig of 15287 bp in length
        *    74600    74699: gap of unknown length
        *    74700    90402: contig of 15703 bp in length
        *    90403    90502: gap of unknown length
        *    90503   112824: contig of 22322 bp in length
        *   112825   112924: gap of unknown length
        *   112925   143717: contig of 30793 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143717
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-757N13"
BASE COUNT    40651 a  29691 c  29709 g  42048 t   1618 others
ORIGIN

Alignment Scores:
Pred. No.:              9.33e-06        Length:         143717
Score:                  114.00          Matches:        20
Percent Similarity:     85.71%          Conservative:   4
Best Local Similarity:  71.43%          Mismatches:     4
Query Match:            59.07%          Indels:         0
DB:                     2               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC069245 (1-143717)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||   |||   :::|||   |||:::|||||||||||||||||||||||||:::||||||
Db  34404 GATGGATTTCCCACCTTGCTGGGTATCCTGGGACCAGAGTATGTAAAACTCATGGGTCTC 34463

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||:::|||||||||   ||||||
Db  34464 TGTATGTGCCTGAGCAGATGCTCT 34487

RESULT 14
AC016472
LOCUS       AC016472              152354 bp    DNA     linear   HTG 04-MAY-2001
DEFINITION  Homo sapiens clone RP11-20P1, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   AC016472
VERSION     AC016472.4  GI:8569170
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152354)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-20P1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152354)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7249022.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3960
                Center clone name: 20_P_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                0.07645259938837792Chemistry: Dye-terminator Big Dye; 100% of
            reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 142312 bases at least Q40
                Consensus quality: 147429 bases at least Q30
                Consensus quality: 149461 bases at least Q20
                Insert size: 153000; agarose-fp
                Insert size: 150754; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality c.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1301: contig of 1301 bp in length
            *     1302 1401: gap of      100 bp
            *     1402     3709: contig of 2308 bp in length
```

```
            *     3710 3809: gap of      100 bp
            *     3810     6730: contig of 2921 bp in length
            *     6731 6830: gap of      100 bp
            *     6831     9206: contig of 2376 bp in length
            *     9207 9306: gap of      100 bp
            *     9307    12163: contig of 2857 bp in length
            *    12164 12263: gap of      100 bp
            *    12264    15381: contig of 3118 bp in length
            *    15382 15481: gap of      100 bp
            *    15482    20100: contig of 4619 bp in length
            *    20101 20200: gap of      100 bp
            *    20201    23885: contig of 3685 bp in length
            *    23886 23985: gap of      100 bp
            *    23986    27226: contig of 3241 bp in length
            *    27227 27326: gap of      100 bp
            *    27327    32869: contig of 5543 bp in length
            *    32870 32969: gap of      100 bp
            *    32970    36903: contig of 3934 bp in length
            *    36904 37003: gap of      100 bp
            *    37004    44369: contig of 7366 bp in length
            *    44370 44469: gap of      100 bp
            *    44470    53277: contig of 8808 bp in length
            *    53278 53377: gap of      100 bp
            *    53378    71089: contig of 17712 bp in length
            *    71090 71189: gap of      100 bp
            *    71190    90983: contig of 19794 bp in length
            *    90984 91083: gap of      100 bp
            *    91084   112306: contig of 21223 bp in length
            *   112307 112406: gap of      100 bp
            *   112407   152354: contig of 39948 bp in length.
FEATURES             Location/Qualifiers
     source          1. .152354
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-20P1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1301
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1402. .3709
                     /note="assembly_fragment"
     misc_feature    3810. .6730
                     /note="assembly_fragment"
     misc_feature    6831. .9206
                     /note="assembly_fragment"
     misc_feature    9307. .12163
                     /note="assembly_fragment"
     misc_feature    12264. .15381
                     /note="assembly_fragment"
     misc_feature    15482. .20100
                     /note="assembly_fragment"
     misc_feature    20201. .23885
                     /note="assembly_fragment"
     misc_feature    23986. .27226
                     /note="assembly_fragment"
     misc_feature    27327. .32869
                     /note="assembly_fragment"
     misc_feature    32970. .36903
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    37004. .44369
                     /note="assembly_fragment"
     misc_feature    44470. .53277
                     /note="assembly_fragment"
     misc_feature    53378. .71089
                     /note="assembly_fragment"
     misc_feature    71190. .90983
                     /note="assembly_fragment"
     misc_feature    91084. .112306
                     /note="assembly_fragment"
     misc_feature    112407. .152354
                     /note="assembly_fragment"
BASE COUNT    42989 a  30395 c  30652 g  46711 t   1607 others
ORIGIN

Alignment Scores:
Pred. No.:              9.89e-06        Length:         152354
Score:                  114.00          Matches:        20
Percent Similarity:     85.71%          Conservative:   4
Best Local Similarity:  71.43%          Mismatches:     4
Query Match:            59.07%          Indels:         0
DB:                     2               Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC016472 (1-152354)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||   :::|||   |||:::||||||||||||||||||||||||:::||||||
Db 117786 GATGGATTTCCCACCTTGCTGGGTATCCTGGGACCAGAGTATGTAAAACTCATGGGTCTC 117845

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||:::|||||||||   ||||||
Db 117846 TGTATGTGCCTGAGCAGATGCTCT 117869

RESULT 15
AC027295
LOCUS       AC027295               160066 bp    DNA     linear   HTG 18-JUN-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-293C20, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC027295
VERSION     AC027295.15  GI:21431081
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160066)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
```

```
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160066)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 160066)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 17, 2002 this sequence version replaced gi:20335550.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HATS
                Center clone name: RP11-293C20
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 42% of reads
                Chemistry: Dye-terminator Big Dye: 58% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 144209 bases at least Q40
                Consensus quality: 149815 bases at least Q30
                Consensus quality: 153559 bases at least Q20
                Estimated insert size: 157473; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3659: contig of 3659 bp in length
            *     3660     3759: gap of unknown length
            *     3760     5914: contig of 2155 bp in length
            *     5915     6014: gap of unknown length
            *     6015     9826: contig of 3812 bp in length
            *     9827     9926: gap of unknown length
            *     9927    12720: contig of 2794 bp in length
            *    12721    12820: gap of unknown length
            *    12821    20586: contig of 7766 bp in length
            *    20587    20686: gap of unknown length
            *    20687    27322: contig of 6636 bp in length
            *    27323    27422: gap of unknown length
            *    27423    39529: contig of 12107 bp in length
            *    39530    39629: gap of unknown length
            *    39630    54210: contig of 14581 bp in length
            *    54211    54310: gap of unknown length
            *    54311    64148: contig of 9838 bp in length
            *    64149    64248: gap of unknown length
            *    64249    79424: contig of 15176 bp in length
            *    79425    79524: gap of unknown length
            *    79525    99543: contig of 20019 bp in length
            *    99544    99643: gap of unknown length
            *    99644   125582: contig of 25939 bp in length
            *   125583   125682: gap of unknown length
            *   125683   160066: contig of 34384 bp in length.
FEATURES            Location/Qualifiers
     source         1. .160066
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="RP11-293C20"
BASE COUNT    47149 a  32911 c  32401 g  46353 t   1252 others
ORIGIN

Alignment Scores:
Pred. No.:                1.04e-05      Length:          160066
Score:                    114.00        Matches:         20
Percent Similarity:       85.71%        Conservative:    4
Best Local Similarity:    71.43%        Mismatches:      4
Query Match:              59.07%        Indels:          0
DB:                       2             Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AC027295 (1-160066)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||   :::|||   |||:::||||||||||||||||||||||||:::|||||||
Db   49872 GATGGATTTCCCACCTTGCTGGGTATCCTGGGACCAGAGTATGTAAAACTCATGGGTCTC 49931

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||:::|||||||||   ||||||
Db   49932 TGTATGTGCCTGAGCAGATGCTCT 49955


Search completed: April 24, 2003, 23:01:33
Job time : 1526 secs
```

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2003, 22:31:13 ; Search time 152 Seconds
(without alignments)
548.184 Million cell updates/sec

Title: US-09-513-999C-7869_COPY_1_37
Perfect score: 193
Sequence: 1 MGGSFALQDSFSSLQGLLGPEYVKLLGLCVCLSGCST 37

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09513999/runat_18042003_170936_28347/app_query.fasta_1.199
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999_@CGN_1_1_200_@runat_18042003_170936_28347 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	193	100.0	447	21	AAC03794	Human secreted pro
	2	193	100.0	447	21	AAZ42680	Human 5' EST isola
c	3	95	49.2	128600	24	ABK83461	Human cDNA differe
c	4	88.5	45.9	570	22	ABA63453	Human foetal liver
c	5	88.5	45.9	570	22	ABA30652	Probe #9118 for ge
c	6	88.5	45.9	570	22	AAK11985	Human brain expres
c	7	88.5	45.9	570	22	AAK37688	Human bone marrow
c	8	88.5	45.9	570	22	AAI18447	Probe #8380 for ge
c	9	88.5	45.9	570	22	AAI43563	Probe #12249 used
c	10	88.5	45.9	570	24	ABS11680	Human genome-deriv
c	11	88	45.6	1982	21	AAC68089	Human secreted pro
c	12	86	44.6	1299	23	AAS91827	DNA encoding novel
c	13	85	44.0	240	24	ABN25478	Human ORFX polynuc
c	14	83	43.0	1909	23	AAS88434	DNA encoding novel
	15	78	40.4	609	22	AAH98980	Human EST-derived
c	16	75	38.9	349	22	AAF65888	Novel human polynu
	17	71	36.8	11534	24	ABL32342	Human immune syste
c	18	69	35.8	2787	23	AAS68019	DNA encoding novel
c	19	69	35.8	2787	23	AAS69828	DNA encoding novel
c	20	67	34.7	1996	22	ABA18896	Human nervous syst
	21	65	33.7	447	22	AAL16101	Human breast cance
	22	65	33.7	447	22	AAL24945	Human breast cance
c	23	64	33.2	36901	20	AAZ23892	Murine LOBO genomi
c	24	64	33.2	38886	20	AAZ23897	Murine LOBO homolo
	25	63	32.6	162	22	ABA75762	Human foetal liver
	26	63	32.6	519	22	ABA63320	Human foetal liver
	27	60.5	31.3	2730	22	AAC84293	Human EXCS encodin
	28	60.5	31.3	24533	22	AAS27689	DNA encoding novel
	29	59	30.6	481	23	ABV05461	Human prostate exp
	30	59	30.6	560	22	AAK78945	Human immune/haema
c	31	58.5	30.3	27666	23	ABL08332	Drosophila melanog
	32	58	30.1	2085	20	AAX01362	Nucleobase permeas
c	33	57.5	29.8	1605	24	ABL40748	Chicken heparanase
	34	57	29.5	2743	22	AAK86888	Human immune/haema
	35	57	29.5	7379	19	AAV49653	Human SC1 DNA. Ho
c	36	56.5	29.3	259	23	ABV05218	Human prostate exp
c	37	56.5	29.3	259	23	ABV14387	Human prostate exp
c	38	56.5	29.3	400	23	ABV35471	Human prostate exp
c	39	56.5	29.3	650	24	ABQ56543	Human colon cancer
	40	56.5	29.3	2076	23	AAS64519	DNA encoding novel
	41	56	29.0	1226	22	AAK53271	Human polynucleoti
	42	56	29.0	1380	24	ABK73394	Bacillus lichenifo
	43	56	29.0	2015	22	AAH57417	Human stomach cell
	44	56	29.0	2073	20	AAX87925	Human protease HPR
c	45	56	29.0	2118	22	AAI60469	Human polynucleoti

ALIGNMENTS

RESULT 1
AAC03794
ID AAC03794 standard; cDNA; 447 BP.
XX
AC AAC03794;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3792.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.

```
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG03788.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 3792; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Alignment Scores:
Pred. No.:              3.61e-20        Length:         447
Score:                  193.00          Matches:        37
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     21              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AAC03794 (1-447)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 2
AAZ42680
ID   AAZ42680 standard; cDNA; 447 BP.
XX
AC   AAZ42680;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human 5' EST isolated from a cDNA library SEQ ID NO:439.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   P-PSDB; AAY65066.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Claim 1; Page 402; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Alignment Scores:
Pred. No.:              3.61e-20        Length:         447
Score:                  193.00          Matches:        37
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     21              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AAZ42680 (1-447)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 3
ABK83461/c
ID   ABK83461 standard; cDNA; 128600 BP.
XX
AC   ABK83461;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #32.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
```

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 32; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
modulating GA; M3 is useful for screening an agent capable of modulating
GCA preferably in an inflammation in a tissue; M4 is useful for
detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 128600 BP; 36139 A; 24970 C; 26316 G; 41175 T; 0 other;

Alignment Scores:

Pred. No.:	0.0459	Length:	128600
Score:	95.00	Matches:	21
Percent Similarity:	64.86%	Conservative:	3
Best Local Similarity:	56.76%	Mismatches:	9
Query Match:	49.22%	Indels:	4
DB:	24	Gaps:	1

US-09-513-999C-7869_COPY_1_37 (1-37) x ABK83461 (1-128600)

```
Qy      3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGln-----------GlyLeuLeu 18
          |||:::               :::   :::|||||||||           |||||||||
Db 126328 GGGGCTATGTGTGGACCTCCCACCTTGACTTCATTGCAGACGCATTTGTTGGGGCTCCTG 126269

Qy     19 GlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
          |||   |||||||||||||||||||||||||   |||||||||   |||
Db 126268 GGGCTGGAGTATGTAAAGCTCCTGGGTCTCTGTGCATGCCTGAGCAGCTGC 126218
```

RESULT 4
ABA63453/c
ID ABA63453 standard; DNA; 570 BP.
XX
AC ABA63453;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #11758.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 11758; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:

Pred. No.:	0.000408	Length:	570
Score:	88.50	Matches:	21
Percent Similarity:	61.11%	Conservative:	1
Best Local Similarity:	58.33%	Mismatches:	7
Query Match:	45.85%	Indels:	7
DB:	22	Gaps:	1

US-09-513-999C-7869_COPY_1_37 (1-37) x ABA63453 (1-570)

```
Qy      1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
          |||   ||||||                          :::|||   |||   ||||||
Db    568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy     21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
          ||||||||||||   ||||||||||||||||||||||||||||   |||
Db    529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482
```

RESULT 5
ABA30652/c
ID ABA30652 standard; DNA; 570 BP.
XX
AC ABA30652;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #9118 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 9118; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:

Pred. No.:	0.000408	Length:	570
Score:	88.50	Matches:	21
Percent Similarity:	61.11%	Conservative:	1
Best Local Similarity:	58.33%	Mismatches:	7
Query Match:	45.85%	Indels:	7
DB:	22	Gaps:	1

US-09-513-999C-7869_COPY_1_37 (1-37) x ABA30652 (1-570)

```
Qy      1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
          |||   ||||||                          :::|||   |||   ||||||
Db    568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy     21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
          ||||||||||||   ||||||||||||||||||||||||||||   |||
Db    529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482
```

RESULT 6
AAK11985/c
ID AAK11985 standard; DNA; 570 BP.
XX
AC AAK11985;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 11976.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 11976; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:

Pred. No.:	0.000408	Length:	570
Score:	88.50	Matches:	21
Percent Similarity:	61.11%	Conservative:	1
Best Local Similarity:	58.33%	Mismatches:	7
Query Match:	45.85%	Indels:	7
DB:	22	Gaps:	1

US-09-513-999C-7869_COPY_1_37 (1-37) x AAK11985 (1-570)

```
Qy      1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
          |||   ||||||                          :::|||   |||   ||||||
Db    568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530
```

```
Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||||||||||||||||||   |||
Db     529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482

RESULT 7
AAK37688/c
ID   AAK37688 standard; DNA; 570 BP.
XX
AC   AAK37688;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 12245.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 12245; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:
Pred. No.:              0.000408        Length:         570
Score:                  88.50           Matches:        21
Percent Similarity:     61.11%          Conservative:   1
Best Local Similarity:  58.33%          Mismatches:     7
Query Match:            45.85%          Indels:         7
DB:                     22              Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x AAK37688 (1-570)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           |||   ||||||                          :::|||   |||   ||||||
Db     568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||||||||||||||||||   |||
Db     529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482

RESULT 8
AAI18447/c
ID   AAI18447 standard; DNA; 570 BP.
XX
AC   AAI18447;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #8380 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 8380; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:
Pred. No.:              0.000408        Length:         570
Score:                  88.50           Matches:        21
Percent Similarity:     61.11%          Conservative:   1
Best Local Similarity:  58.33%          Mismatches:     7
Query Match:            45.85%          Indels:         7
DB:                     22              Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x AAI18447 (1-570)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           |||   ||||||                          :::|||   |||   ||||||
Db     568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||||||||||||||||||   |||
Db     529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482

RESULT 9
AAI43563/c
ID   AAI43563 standard; DNA; 570 BP.
```

XX
AC AAI43563;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #12249 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 12249; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:

Pred. No.:	0.000408	Length:	570
Score:	88.50	Matches:	21
Percent Similarity:	61.11%	Conservative:	1
Best Local Similarity:	58.33%	Mismatches:	7
Query Match:	45.85%	Indels:	7
DB:	22	Gaps:	1

US-09-513-999C-7869_COPY_1_37 (1-37) x AAI43563 (1-570)

```
Qy      1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
          |||   ||||||                        :::|||   |||   ||||||
Db    568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy     21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
          ||||||||||||   |||||||||||||||||||||||||||||   |||
Db    529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482
```

RESULT 10
ABS11680/c
ID ABS11680 standard; DNA; 570 BP.
XX
AC ABS11680;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID No 11671.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 1; SEQ ID No 11671; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe of the invention.

```
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Alignment Scores:
Pred. No.:              0.000408      Length:          570
Score:                  88.50         Matches:         21
Percent Similarity:     61.11%        Conservative:    1
Best Local Similarity:  58.33%        Mismatches:      7
Query Match:            45.85%        Indels:          7
DB:                     24            Gaps:            1

US-09-513-999C-7869_COPY_1_37 (1-37) x ABS11680 (1-570)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           |||   ||||||                          :::|||   |||   ||||||
Db     568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||||||||||||||||||   |||
Db     529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482

RESULT 11
AAC68089/c
ID   AAC68089 standard; cDNA; 1982 BP.
XX
AC   AAC68089;
XX
DT   20-FEB-2001  (first entry)
XX
DE   Human secreted protein cDNA sequence #9.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058335-A1.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US07534.

     26-MAR-1999;   99US-0126598.
     22-DEC-1999;   99US-0171504.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-611702/58.
DR   P-PSDB; AAB37356.
XX
PT   Nucleic acids encoding human secreted proteins, used to treat, prevent,
PT   ameliorate or diagnose conditions such as cancer, and autoimmune
PT   diseases e.g. arthritis -
XX
PS   Claim 1; Pages 321-322; 387pp; English.
XX
CC   The invention relates to the isolation of genes AAC68081-C68127 encoding
CC   47 human secreted proteins AAB37348-B37394. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (AAC68072) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 1982 BP; 549 A; 513 C; 471 G; 448 T; 1 other;

Alignment Scores:
Pred. No.:              0.00243       Length:          1982
Score:                  88.00         Matches:         20
Percent Similarity:     67.74%        Conservative:    1
Best Local Similarity:  64.52%        Mismatches:      10
Query Match:            45.60%        Indels:          0
DB:                     21            Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AAC68089 (1-1982)

Qy       6 AlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeu 25
           |||||||||:::          |||   |||      ||||||||||||||||||||
Db    1156 GCTTTGCAGGAGTTGCAGTTACTTTTGCTGGGAAACCCAGGGCCGGAGTATGTAAAGCTG 1097

Qy      26 LeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   |||   ||||||   |||
Db    1096 CTGGGTCTCTGCACGTGCTCAAGCGGCTACTCT 1064

RESULT 12
AAS91827/c
ID   AAS91827 standard; cDNA; 1299 BP.
XX
AC   AAS91827;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27631.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27640.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27631; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
```

```
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1299 BP; 399 A; 349 C; 271 G; 280 T; 0 other;

Alignment Scores:
Pred. No.:                0.00284        Length:          1299
Score:                    86.00          Matches:         21
Percent Similarity:       62.16%         Conservative:    2
Best Local Similarity:    56.76%         Mismatches:      12
Query Match:              44.56%         Indels:          2
DB:                       23             Gaps:            1

US-09-513-999C-7869_COPY_1_37 (1-37) x AAS91827 (1-1299)

Qy       2 GlyGlySerPheAla------LeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGly 19
           |||||||||   |||      |||   :::            |||   |||:::||||||
Db     916 GGCGGCAGCGAGGCTCCCTCCTTGCCTGAGTTGCAGCCACATTTGTTGGGGATCCTGGGG 857

Qy      20 ProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
             ||||||||||||||   |||||||||   |||   |||   ||||||
Db     856 CTGGAGTATGTAAAGCTCCCAGGTCTTTGTGCATGCCGGAGCAGCTGCTCT 806

RESULT 13
ABN25478/c
ID   ABN25478 standard; cDNA; 240 BP.
XX
AC   ABN25478;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:19433.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP09726.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders
XX
PS   Disclosure; SEQ ID 19433; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 240 BP; 64 A; 61 C; 61 G; 54 T; 0 other;

Alignment Scores:
Pred. No.:                0.000455       Length:          240
Score:                    85.00          Matches:         14
Percent Similarity:       94.12%         Conservative:    2
Best Local Similarity:    82.35%         Mismatches:      1
Query Match:              44.04%         Indels:          0
DB:                       24             Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x ABN25478 (1-240)

Qy      20 ProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           |||||||||:::||||||||||||||||||:::|||||||||   ||||||
Db     150 CCCGAGTATCTAAAGCTCCTGGGGCTCTGTATATGCCTGAGTAGCTGCTCT 100

RESULT 14
AAS88434/c
ID   AAS88434 standard; cDNA; 1909 BP.
XX
AC   AAS88434;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #24238.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24247.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24238; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1909 BP; 705 A; 443 C; 389 G; 372 T; 0 other;

Alignment Scores:

Pred. No.:	0.0134	Length:	1909
Score:	83.00	Matches:	18
Percent Similarity:	60.61%	Conservative:	2
Best Local Similarity:	54.55%	Mismatches:	13
Query Match:	43.01%	Indels:	0
DB:	23	Gaps:	0

09-513-999C-7869_COPY_1_37 (1-37) x AAS88434 (1-1909)

```
Qy       4 SerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrVal 23
           |||      |||   :::      |||   |||   |||         |||||||||:::
Db     205 TCAAGGACCTTGCCTGAGTTACAGTCCCATTTGCCAGGGAACCGAGAGCCAGAGTACATA 146

Qy      24 LysLeuGluGlyLeuCysValCysLeuSerGlyCysSer 36
           |||   ||||||||||||   |||||||||   ||||||
Db     145 AAGCCCCTGGGTCTTTGTGCATGCCTGAGCAGCTGCTCC 107
```

RESULT 15
AAH98980
ID AAH98980 standard; cDNA; 609 BP.
XX
AC AAH98980;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 837.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM24321.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 694; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 609 BP; 99 A; 162 C; 173 G; 175 T; 0 other;

Alignment Scores:

Pred. No.:	0.0176	Length:	609
Score:	78.00	Matches:	17
Percent Similarity:	71.43%	Conservative:	3
Best Local Similarity:	60.71%	Mismatches:	8
Query Match:	40.41%	Indels:	0
DB:	22	Gaps:	0

US-09-513-999C-7869_COPY_1_37 (1-37) x AAH98980 (1-609)

```
Qy       6 AlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeu 25
           ||||||   :::      |||         |||:::|||      ||||||||||||||||
Db     273 GCCTTGCCTGAGATGCCATCACCTTTCTTGGGGATCCTAAGGCTGGAGTATGTAAAGCTC 332

Qy      26 LeuGlyLeuCysValCysLeuSer 33
           ||||||||||||:::|||||||||
Db     333 CTGGGTCTCTGTATGTGCCTGAGC 356
```

Search completed: April 24, 2003, 23:01:50
Job time : 165 secs

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2003, 22:34:13 ; Search time 1054 Seconds
(without alignments)
568.532 Million cell updates/sec

Title: US-09-513-999C-7869_COPY_1_37
Perfect score: 193
Sequence: 1 MGGSFALQDSFSSLQGLLGPEYVKLLGLCVCLSGCST 37

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

rched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09513999/runat_18042003_170937_28370/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999_@CGN_1_1_1456_@runat_18042003_170937_28370 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	193	100.0	470	17	AQ770688	AQ770688 HS_5368_B
2	169	87.6	525	17	AQ165256	AQ165256 HS_3025_B
c 3	129	66.8	864	17	AQ739814	AQ739814 HS_5505_A
c 4	109	56.5	519	17	AQ186743	AQ186743 HS_3109_B
5	109	56.5	533	17	AQ358845	AQ358845 HS_5028_B
6	108	56.0	628	17	AQ237815	AQ237815 RPCI11-70
7	105	54.4	292	17	AQ508480	AQ508480 RPCI-11-2
c 8	105	54.4	632	17	AG146082	AG146082 Pan trogl
9	104	53.9	666	17	AQ427698	AQ427698 CITBI-E1-
c 10	102	52.8	418	9	AA493535	AA493535 ng75g10.s
c 11	102	52.8	556	17	AQ384817	AQ384817 RPCI11-13
12	102	52.8	695	17	AG179297	AG179297 Pan trogl
c 13	101	52.3	703	17	AQ534396	AQ534396 RPCI-11-3
14	100	51.8	634	17	AG160901	AG160901 Pan trogl
15	100	51.8	659	17	AG151043	AG151043 Pan trogl
c 16	100	51.8	676	17	AG061401	AG061401 Pan trogl
17	100	51.8	769	17	AQ899390	AQ899390 HS_5234_A
c 18	99	51.3	362	17	AQ102366	AQ102366 HS_3040_A
19	99	51.3	410	17	AQ442274	AQ442274 HS_5137_A
20	99	51.3	416	17	AQ182486	AQ182486 HS_3077_A
21	99	51.3	453	17	AQ437684	AQ437684 HS_5137_A
c 22	99	51.3	485	17	AQ671849	AQ671849 HS_5462_A
c 23	99	51.3	635	17	AQ390599	AQ390599 CITBI-E1-
24	98	50.8	546	17	AQ435071	AQ435071 HS_5114_B
25	98	50.8	553	17	AZ521751	AZ521751 RPCI-11-1
c 26	97	50.3	597	17	AQ506984	AQ506984 RPCI-11-3
c 27	96.5	50.0	410	17	AQ531865	AQ531865 RPCI-11-3
c 28	96.5	50.0	672	17	AQ551272	AQ551272 RPCI-11-4
c 29	96	49.7	347	17	AQ631315	AQ631315 RPCI-11-4
c 30	96	49.7	444	17	AQ463109	AQ463109 HS_5211_A
31	96	49.7	683	17	AG091225	AG091225 Pan trogl
32	95	49.2	360	17	AQ207172	AQ207172 HS_3239_B
c 33	95	49.2	519	17	AQ139984	AQ139984 HS_3106_A
34	95	49.2	546	17	AQ541696	AQ541696 RPCI-11-3
35	95	49.2	672	17	AG051939	AG051939 Pan trogl
c 36	94	48.7	363	17	AQ120796	AQ120796 HS_3076_A
c 37	94	48.7	425	17	AQ683450	AQ683450 HS_5432_B
38	94	48.7	452	17	AQ534129	AQ534129 RPCI-11-3
39	94	48.7	516	17	AQ457001	AQ457001 HS_5151_A
c 40	94	48.7	546	17	AQ333597	AQ333597 HS_5008_A
c 41	93	48.2	529	17	AQ881246	AQ881246 HS_5137_B
c 42	92	47.7	580	17	AQ532835	AQ532835 RPCI-11-3
c 43	91	47.2	463	17	AQ550196	AQ550196 RPCI-11-4
44	91	47.2	546	17	AQ468922	AQ468922 HS_5139_A
c 45	91	47.2	559	17	AQ385332	AQ385332 RPCI11-14

ALIGNMENTS

RESULT 1
AQ770688
LOCUS AQ770688 470 bp DNA linear GSS 28-JUL-1999
DEFINITION HS_5368_B2_C08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=944 Col=16 Row=F, DNA sequence.
ACCESSION AQ770688
VERSION AQ770688.1 GI:5648804
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

```
          Hood,L.
  TITLE   Sequence-tagged connectors: A sequence approach to mapping and
          scanning the human genome
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE 99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
          High Throughput Sequencing Center
          University of Washington
          401 Queen Anne Avenue North, Seattle, WA 98109, USA
          Tel: (206) 616-3618
          Fax: (206) 616-3887
          Email: jwallace@u.washington.edu
          Clones are derived from the human BAC library RPCI-11. For BAC
          library availability, please contact Pieter de Jong
          (pieter@dejong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
          or from Resear h Genetics (info@resgen.com). BAC end Web Server:
          http://www.htsc.washington.edu
          Plate: 944  row: F  column: 16
          Seq primer: SP6
          Class: BAC ends
          High quality sequence stop: 470.
FEATURES            Location/Qualifiers
     source         1. .470
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate=944 Col=16 Row=F"
                    /clone_lib="RPCI-11 Human Male BAC Library"
                    /sex="male"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    Male blood DNA was isolated from one randomly chosen donor
                    and partially digested with a combination of EcoRI and
                    EcoRI Methylase. Size selected DNA was cloned into the
                    pBACe3.6 vector at EcoRI sites"
BASE COUNT       83 a    112 c    131 g    141 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                4.59e-16       Length:         470
Score:                    193.00         Matches:        37
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       17             Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ770688 (1-470)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 162

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 213

RESULT 2
AQ165256
LOCUS       AQ165256                525 bp    DNA     linear   GSS 16-OCT-1998
DEFINITION  HS_3025_B2_G06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=12 Row=N, DNA sequence.
ACCESSION   AQ165256
VERSION     AQ165256.1  GI:3563451
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3025  row: N  column: 12
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES              Location/Qualifiers
     source           1. .525
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3025 Col=12 Row=N"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      102 a    139 c    137 g    143 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:                8.15e-13       Length:         525
Score:                    169.00         Matches:        33
Percent Similarity:       91.89%         Conservative:   1
Best Local Similarity:    89.19%         Mismatches:     3
Query Match:              87.56%         Indels:         0
DB:                       17             Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ165256 (1-525)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           |||||||||||||||||||||||:::      ||||||||||||||||||||||||||||
Db      66 ATGGGCGGATCTTTCGCCTTGCAGGAAATTATNTCATCTTTGCAGGGACTTCTGGGGCCG 125

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||||||||||||||||   ||||||||||||||||||||||||||||||
Db     126 GAGTATGTAAAACTCCTGNGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 176

RESULT 3
AQ739814/c
LOCUS       AQ739814                864 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_5505_A1_A09_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1081 Col=17 Row=A, DNA sequence.
ACCESSION   AQ739814
VERSION     AQ739814.1  GI:5517336
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
```

```
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1081  row: A  column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 864.
FEATURES              Location/Qualifiers
     source           1. .864
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=1081 Col=17 Row=A"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      247 a    182 c    178 g    257 t
ORIGIN

Alignment Scores:
Pred. No.:              3.03e-07       Length:         864
Score:                  129.00         Matches:        23
Percent Similarity:     92.86%         Conservative:   3
Best Local Similarity:  82.14%         Mismatches:     2
Query Match:            66.84%         Indels:         0
DB:                     17             Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ739814 (1-864)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   ||||||:::|||   |||:::||||||||:::||||||||||||||||||||
Db     585 GACAAATTTTCCACCTTGGCAGGGGTCCTGGGTCCACAGTATGTAAAACTCCTGGGCCTC 526

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||||||||||||||||||||||
Db     525 TGCGTGTGCCTGAGTGGCTGCTCT 502

RESULT 4
AQ186743/c
LOCUS       AQ186743                519 bp    DNA     linear   GSS 01-NOV-1998
DEFINITION  HS_3109_B1_A06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3109 Col=11 Row=B, DNA sequence.
ACCESSION   AQ186743
VERSION     AQ186743.1  GI:3586185
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3109  row: B  column: 11
            Class: BAC ends
            High quality sequence stop: 519.
FEATURES              Location/Qualifiers
     source           1. .519
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3109 Col=11 Row=B"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      144 a    155 c    122 g     93 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              7.79e-05       Length:         519
Score:                  109.00         Matches:        21
Percent Similarity:     82.76%         Conservative:   3
Best Local Similarity:  72.41%         Mismatches:     5
Query Match:            56.48%         Indels:         0
DB:                     17             Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ186743 (1-519)

Qy       8 GlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGly 27
           :::|||      |||:::|||   |||:::||||||   ||||||||||||||||||||
Db     399 AAGGATGGATCTTCTGCCTTGGCGGGAATCCTGGGGCTGGAGTATGTAAAACTCTTGGGT 340

Qy      28 LeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||
Db     339 CTCTGTGTGTGCCCCAGTGGCTGCTCT 313

RESULT 5
AQ358845
LOCUS       AQ358845                533 bp    DNA     linear   GSS 06-MAR-1999
DEFINITION  HS_5028_B1_B02_T7 RPCI11 Human Male BAC Library Homo sapiens
            genomic clone Plate=604 Col=3 Row=D, DNA sequence.
ACCESSION   AQ358845
VERSION     AQ358845.1  GI:4207721
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 604  row: D  column: 3
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 533.
FEATURES              Location/Qualifiers
     source           1. .533
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=604 Col=3 Row=D"
                      /clone_lib="RPCI11 Human Male BAC Library"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT       98 a    132 c    139 g    151 t     13 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:              8.03e-05        Length:         533
Score:                  109.00          Matches:        19
Percent Similarity:     82.14%          Conservative:   4
Best Local Similarity:  67.86%          Mismatches:     5
Query Match:            56.48%          Indels:         0
DB:                     17              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ358845 (1-533)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||   :::|||   |||:::|||||||||||||||   ||||||:::|||||||
Db     132 GATGGATTTCCCACCTTGCTGGGTATCCTGGGACCAGAGTATTGTAAACTCATGGGTCTC 191

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||:::|||||||||   ||||||
Db     192 TGTATGTGCCTGAGCAGATGCTCT 215

RESULT 6
AQ237815
LOCUS       AQ237815                628 bp    DNA     linear   GSS 21-APR-1999
DEFINITION  RPCI11-70H4.TK RPCI-11 Homo sapiens genomic clone RPCI-11-70H4, DNA
            sequence.
ACCESSION   AQ237815
VERSION     AQ237815.1  GI:3670106
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Homo sapiens"
                     /db_xref="GDB:7526667"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-70H4"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      125 a    150 c    164 g    189 t
ORIGIN

Alignment Scores:
Pred. No.:              0.000132        Length:         628
Score:                  108.00          Matches:        20
Percent Similarity:     91.67%          Conservative:   2
Best Local Similarity:  83.33%          Mismatches:     2
Query Match:            55.96%          Indels:         0
DB:                     17              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ237815 (1-628)

Qy      14 LeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSer 33
           |||      :::|||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 TTGCCGCAGGTCCTGGGGCCAGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGC 195

Qy      34 GlyCysSerThr 37
           |||||||||:::
Db     196 GGCTGCTCTTCA 207

RESULT 7
AQ508480
LOCUS       AQ508480                292 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  RPCI-11-294M24.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-294M24
            , DNA sequence.
ACCESSION   AQ508480
VERSION     AQ508480.1  GI:4713227
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-294M24.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="GDB:7612823"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-294M24"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       71 a     69 c     78 g     73 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              0.000138        Length:         292
Score:                  105.00          Matches:        19
Percent Similarity:     91.67%          Conservative:   3
Best Local Similarity:  79.17%          Mismatches:     2
Query Match:            54.40%          Indels:         0
DB:                     17              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ508480 (1-292)

Qy      13 SerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeu 32
           :::||||||||||:::   ||||||||||||:::|||||||||   ||||||||||||||||
Db      81 GCCTTGCAGGGGATCTCAGGGCCAGAGTATCTAAAACTCCTGGTTCTCTGTGTGTGCCTA 140

Qy      33 SerGlyCysSer 36
           ||||||||||||
```

```
Db    141 AGTGGCTGCTCT 152

RESULT 8
AG146082/c
LOCUS       AG146082                632 bp    DNA     linear   GSS 08-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-007J05.TJ, genomic survey
            sequence.
ACCESSION   AG146082
VERSION     AG146082.1  GI:16675760
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-007J05.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 632)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-007J05.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      186 a    160 c    156 g    130 t
ORIGIN

Alignment Scores:
Pred. No.:                0.000332        Length:          632
Score:                    105.00          Matches:         19
Percent Similarity:       82.14%          Conservative:    4
Best Local Similarity:    67.86%          Mismatches:      5
Query Match:              54.40%          Indels:          0
DB:                       17              Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AG146082 (1-632)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||      |||:::|||      :::|||||||||||||||:::|||||||||||
Db     392 GATGTATCTTCCACTTTGCCAAGGATCCTGGGGCCAGAGTATTTAAAACTTCTTGGTTTT 333

Qy      29 CysValCysLeuSerGlyCysSer 36
           ||||||||||||:::|||||||||
Db     332 TGTGTGTGTCTGAATGGATGCTCT 309

RESULT 9
AQ427698
LOCUS       AQ427698                666 bp    DNA     linear   GSS 24-MAR-1999
DEFINITION  CITBI-E1-2575E9.TR CITBI-E1 Homo sapiens genomic clone 2575E9, DNA
            sequence.
ACCESSION   AQ427698
VERSION     AQ427698.1  GI:4500605
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2575E9.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2575E9"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      136 a    157 c    171 g    201 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                0.000479        Length:          666
Score:                    104.00          Matches:         22
Percent Similarity:       76.47%          Conservative:    4
Best Local Similarity:    64.71%          Mismatches:      8
Query Match:              53.89%          Indels:          0
DB:                       17              Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ427698 (1-666)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
             |||      ||||||   :::      :::|||||||   |||:::|||||||||||:::
Db     365 GGACTTCACGCCCTGCCTGAGTTGCAGACAAGTTTGTTGGGGATCCTTGGGCCAGAGTTT 424

Qy      23 ValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||||||||||||||   ||||||   ||||||
Db     425 GTAAAGCTCCTGGGTCTCTGTGTGTGACTGAGCAGCTGCTCT 466

RESULT 10
AA493535/c
LOCUS       AA493535                418 bp    mRNA    linear   EST 18-AUG-1997
DEFINITION  ng75g10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:940674
            similar to contains element PTR7 repetitive element ;, mRNA
            sequence.
ACCESSION   AA493535
VERSION     AA493535.1  GI:2223376
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

```
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 529    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:940674"
                     /clone_lib="NCI_CGAP_Pr6"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from prostatic
                     intraepithelial neoplasia (low-grade), cDNA  made by
                     oligo-dT priming.  Non-directionally cloned.
                     Size-selected  on agarose gel, average insert size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT      106 a    127 c     96 g     89 t
ORIGIN

Alignment Scores:
Pred. No.:                0.00052        Length:         418
Score:                    102.00         Matches:        20
Percent Similarity:       75.00%         Conservative:   1
Best Local Similarity:    71.43%         Mismatches:     7
Query Match:              52.85%         Indels:         0
DB:                       9              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AA493535 (1-418)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||       |||   |||:::      ||||||||||||||||||||||||||
Db     281 GACAGATTTCTTGTCTTGCCAGGGATCCCAGAGCCGGAGTATGTAAAACTCTTGGGTCTC 222

Qy      29 CysValCysLeuSerGlyCysSer 36
           |||||||||||||||   ||||||
Db     221 TGTGTATGCCTGAGCAGCTGCTCT 198

RESULT 11
AQ384817/c
LOCUS       AQ384817                 556 bp    DNA     linear   GSS 21-MAY-1999
DEFINITION  RPCI11-138P7.TV RPCI-11 Homo sapiens genomic clone RPCI-11-138P7,
            DNA sequence.
ACCESSION   AQ384817
VERSION     AQ384817.1  GI:4355840
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-138P7.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /db_xref="GDB:7552974"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-138P7"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      151 a    141 c    142 g    121 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                0.00072        Length:         556
Score:                    102.00         Matches:        19
Percent Similarity:       80.77%         Conservative:   2
Best Local Similarity:    73.08%         Mismatches:     5
Query Match:              52.85%         Indels:         0
DB:                       17             Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ384817 (1-556)

Qy      11 PheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysVal 30
           |||   :::|||   |||:::   ||||||||||||||||||||||      ||||||
Db     349 TTTCCTGCCTTGCTAGGGATCCCAGGGCCAGAGTATGTAAAACTCCTACGTTTCTGTGTG 290

Qy      31 CysLeuSerGlyCysSer 36
           ||||||||||||||||||
Db     289 TGCTTGAGTGGCTGCTCT 272

RESULT 12
AG179297
LOCUS       AG179297                 695 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-051I11.TJ, genomic survey
            sequence.
ACCESSION   AG179297
VERSION     AG179297.1  GI:16708977
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-051I11.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 695)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
```

```
            PRIMERS
                Sequencing: TJ
            LIBRARY
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES            Location/Qualifiers
     source         1. .695
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="RP43-051I11.TJ"
                    /sex="male"
                    /cell_type="lymphocytes"
                    /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      243 a    135 c    122 g    192 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              0.000929        Length:         695
Score:                  102.00          Matches:        19
Percent Similarity:     78.57%          Conservative:   3
Best Local Similarity:  67.86%          Mismatches:     6
Query Match:            52.85%          Indels:         0
DB:                     17              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AG179297 (1-695)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||   |||   :::      |||::::::   ||||||||||||||||||||||||||
Db    135 GACAGATTTCCCACCTTTCCAGGGATCATGCGGCCAGAGTATGTAAAACTCCTGGGTCTC 194

Qy     29 CysValCysLeuSerGlyCysSer 36
          ||||||   ||||||||||||||||
Db    195 TGTGTGTATCTGAGTGGCTGCTCT 218

RESULT 13
AQ534396/c
LOCUS       AQ534396                703 bp    DNA     linear   GSS 18-MAY-1999
DEFINITION  RPCI-11-380I3.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-380I3,
            DNA sequence.
ACCESSION   AQ534396
VERSION     AQ534396.1  GI:4846086
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .703
                    /organism="Homo sapiens"
                    /db_xref="GDB:7645730"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-380I3"
                    /clone_lib="RPCI-11"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
BASE COUNT      201 a    194 c    165 g    143 t
ORIGIN

Alignment Scores:
Pred. No.:              0.00128         Length:         703
Score:                  101.00          Matches:        17
Percent Similarity:     75.00%          Conservative:   4
Best Local Similarity:  60.71%          Mismatches:     7
Query Match:            52.33%          Indels:         0
DB:                     17              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x AQ534396 (1-703)

Qy      9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
          |||   :::      |||   |||:::||||||||||||||||:::|||   ||||||
Db    389 GATGGATATCCCCCCTTGCTGGGGATCTTGGGGCCAGAGTATATAAAATTCCTGGGTCAC 330

Qy     29 CysValCysLeuSerGlyCysSer 36
          |||:::||||||||   ||||||
Db    329 TGTATGTGCCTGAGCAGCTGCTCT 306

RESULT 14
AG160901
LOCUS       AG160901                634 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-026N23.T7, genomic survey
            sequence.
ACCESSION   AG160901
VERSION     AG160901.1  GI:16690579
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-026N23.T7.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 634)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: T7
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES            Location/Qualifiers
     source         1. .634
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="RP43-026N23.T7"
                    /sex="male"
                    /cell_type="lymphocytes"
                    /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      116 a    168 c    156 g    194 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.00154        Length:          634
Score:                    100.00         Matches:         19
Percent Similarity:       82.14%         Conservative:    4
Best Local Similarity:    67.86%         Mismatches:      5
Query Match:              51.81%         Indels:          0
DB:                       17             Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AG160901 (1-634)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||         :::|||:::|||:::||||||||||:::||||||   ||||||||||||
Db     115 GACAGATCTCCTGCCTTAAAGGGGATCCTGGGGCCACAATATGTAAATCTCCTGGGTCTC 174

Qy      29 CysValCysLeuSerGlyCysSer 36
           ||||||   ||||||||||||||||
Db     175 TGTGTGTGACTGAGTGGCTGCTCT 198

RESULT 15
AG151043
LOCUS       AG151043                659 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-014A14.T7, genomic survey
            sequence.
ACCESSION   AG151043
VERSION     AG151043.1  GI:16680721
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-014A14.T7.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 659)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: T7
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-014A14.T7"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      129 a    159 c    154 g    217 t
ORIGIN

Alignment Scores:
Pred. No.:                0.00161        Length:          659
Score:                    100.00         Matches:         19
Percent Similarity:       82.14%         Conservative:    4
Best Local Similarity:    67.86%         Mismatches:      5
Query Match:              51.81%         Indels:          0
DB:                       17             Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x AG151043 (1-659)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||         :::|||:::|||:::||||||||||:::||||||   ||||||||||||
Db     116 GACAGATCTCCTGCCTTAAAGGGGATCCTGGGGCCACAATATGTAAATCTCCTGGGTCTC 175

Qy      29 CysValCysLeuSerGlyCysSer 36
           ||||||   ||||||||||||||||
Db     176 TGTGTGTGACTGAGTGGCTGCTCT 199


Search completed: April 24, 2003, 23:19:32
Job time : 1058 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2003, 22:35:13 ; Search time 42 Seconds
(without alignments)
270.168 Million cell updates/sec

Title: US-09-513-999C-7869_COPY_1_37
Perfect score: 193
Sequence: 1 MGGSFALQDSFSSLQGLLGPEYVKLLGLCVCLSGCST 37

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09513999/runat_18042003_170937_28379/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999_@CGN_1_1_32_@runat_18042003_170937_28379 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	58	30.1	2085	2	US-08-677-049-1	Sequence 1, Appli
	2	57	29.5	7379	4	US-09-341-587-5	Sequence 5, Appli
	3	56	29.0	2073	4	US-09-032-523-6	Sequence 6, Appli
	4	56	29.0	3120	1	US-08-456-647B-19	Sequence 19, Appl
	5	56	29.0	3120	2	US-08-237-401A-19	Sequence 19, Appl
c	6	54.5	28.2	2064	3	US-08-875-944B-1	Sequence 1, Appli
c	7	54.5	28.2	2064	4	US-09-116-049-3	Sequence 3, Appli
c	8	54.5	28.2	2410	2	US-08-780-835B-1	Sequence 1, Appli
c	9	54.5	28.2	2410	4	US-09-303-268-1	Sequence 1, Appli
c	10	54.5	28.2	2410	4	US-09-116-049-1	Sequence 1, Appli
c	11	54	28.0	169998	4	US-09-676-610B-24	Sequence 24, Appl
c	12	53.5	27.7	1308	4	US-09-526-993-4	Sequence 4, Appli
c	13	53.5	27.7	2167	4	US-09-526-993-3	Sequence 3, Appli
c	14	53.5	27.7	2176	1	US-07-778-890A-2	Sequence 2, Appli
c	15	53.5	27.7	3224	4	US-09-526-993-2	Sequence 2, Appli
c	16	53.5	27.7	6070	4	US-09-526-993-1	Sequence 1, Appli
c	17	53.5	27.7	6157	4	US-09-526-993-10	Sequence 10, Appl
c	18	53.5	27.7	6202	4	US-09-526-993-8	Sequence 8, Appli
c	19	52.5	27.2	59065	4	US-09-813-817-3	Sequence 3, Appli
c	20	52.5	27.2	59065	4	US-09-978-197-3	Sequence 3, Appli
c	21	52	26.9	2055	3	US-08-872-855-3	Sequence 3, Appli
c	22	52	26.9	2800	3	US-08-872-855-1	Sequence 1, Appli
	23	51.5	26.7	1019	4	US-09-177-650-128	Sequence 128, App
	24	51.5	26.7	1191	2	US-09-061-337-11	Sequence 11, Appl
	25	51.5	26.7	1191	2	US-09-122-129-11	Sequence 11, Appl
	26	51.5	26.7	1191	3	US-09-340-991-11	Sequence 11, Appl
	27	51.5	26.7	1191	3	US-08-974-609-11	Sequence 11, Appl
	28	51.5	26.7	1191	4	US-09-549-098-11	Sequence 11, Appl
c	29	51.5	26.7	1297	3	US-09-083-521-4	Sequence 4, Appli
	30	51.5	26.7	3523	3	US-08-749-527-1	Sequence 1, Appli
	31	51	26.4	3390	1	US-08-453-742-26	Sequence 26, Appl
	32	51	26.4	3390	1	US-08-454-464-26	Sequence 26, Appl
	33	51	26.4	3390	1	US-08-453-222-26	Sequence 26, Appl
	34	51	26.4	3390	1	US-08-452-802-26	Sequence 26, Appl
	35	51	26.4	3416	1	US-08-453-742-24	Sequence 24, Appl
	36	51	26.4	3416	1	US-08-454-464-24	Sequence 24, Appl
	37	51	26.4	3416	1	US-08-453-222-24	Sequence 24, Appl
	38	51	26.4	3416	1	US-08-452-802-24	Sequence 24, Appl
	39	51	26.4	3588	1	US-07-792-885A-2	Sequence 2, Appli
	40	51	26.4	9370	1	US-08-320-559-27	Sequence 27, Appl
	41	51	26.4	9370	3	US-08-545-860D-27	Sequence 27, Appl
	42	51	26.4	9370	5	PCT-US94-04496-27	Sequence 27, Appl
	43	51	26.4	9391	1	US-08-320-559-25	Sequence 25, Appl
	44	51	26.4	9391	3	US-08-545-860D-25	Sequence 25, Appl
	45	51	26.4	9391	5	PCT-US94-04496-25	Sequence 25, Appl

ALIGNMENTS

```
RESULT 1
US-08-677-049-1
; Sequence 1, Application US/08677049
; Patent No. 5858707
;  GENERAL INFORMATION:
;    APPLICANT:  Guimaraes, M. Jorge
;    APPLICANT:  Bazan, J. Fernando
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Zlotnik, Albert
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
;    TITLE OF INVENTION:  NUCLEIC ACIDS; ANTIBODIES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,049
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/000,788
;      FILING DATE:  03-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0511
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  415-852-9196
;       TELEFAX:  415-496-1200
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2085 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  48..1883
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  366..1883
;       OTHER INFORMATION:  /note= "Form 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
        LOCATION:  1515..1883
        OTHER INFORMATION:  /note= "Form 3"
      FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1659..1883
;       OTHER INFORMATION:  /note= "Form 4"
US-08-677-049-1

Alignment Scores:
Pred. No.:              19.6          Length:        2085
Score:                  58.00         Matches:       20
Percent Similarity:     41.07%        Conservative:  3
Best Local Similarity:  35.71%        Mismatches:    13
Query Match:            30.05%        Indels:        20
DB:                     2             Gaps:          3

US-09-513-999C-7869_COPY_1_37 (1-37) x US-08-677-049-1 (1-2085)

Qy       2 GlyGlySerPheAlaLeuGln----------------------AspSerPheSerSer 13
           ||||||         ||||||                      |||||||||
Db     799 GGAGGCCATCTTCAACTTCAACTCACATTTGTATTCCCGTCTTCCGACTCCTTTCGGTGC 858

Qy      14 LeuGln--------------GlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||                 |||         |||   :::|||::::::|||   |||
Db     859 TTGCCCCTGTGGCCTGTGTGTGGTTCATCTCTGCCTTTGTGGGTACGAGTGTTATCCCTC 918

Qy      29 CysValCysLeuSer--------------------GlyCysSerThr 37
           |||   |||||||||                    ||||||   |||
Db     919 TGCAGCTGTCTGAGCCCTCGGATGCACCTTGGTTTTGGCTGCCACACC 966

RESULT 2
US-09-341-587-5
; Sequence 5, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
;  APPLICANT: Mollenhauer, Jan
;  TITLE OF INVENTION: Protein Containing an SRCR Domain
;  FILE REFERENCE: 4121-108
;  CURRENT APPLICATION NUMBER: US/09/341,587
;  CURRENT FILING DATE: 1999-08-31
;  EARLIER APPLICATION NUMBER: PCT/DE98/00096
;  EARLIER FILING DATE: 1998-01-09
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 7379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-341-587-5

Alignment Scores:
Pred. No.:              143           Length:        7379
Score:                  57.00         Matches:       17
Percent Similarity:     57.58%        Conservative:  2
Best Local Similarity:  51.52%        Mismatches:    10
Query Match:            29.53%        Indels:        4
DB:                     4             Gaps:          1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-341-587-5 (1-7379)

Qy       5 PheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLys 24
           |||:::|||      |||||||||            |||||||||      |||
Db    2611 TTCTCTCTCCATCTCTCTTTCTCT-----------TTATTGGGCAGCTATTATCACCTT 2658

Qy      25 LeuLeuGlyLeuCysValCysLeuSerGlyCysSerThr 37
             |||   ||||||:::|||||||||      ||||||
Db    2659 TCTCTCTGTCTCTGTCTCTGTCTCTCTTTCTCTTCAACT 2697

RESULT 3
US-09-032-523-6
; Sequence 6, Application US/09032523
; Patent No. 6232454
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl
;    APPLICANT:  Baugh, Mariah
;    TITLE OF INVENTION:  HUMAN PROTEINASE MOLECULES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/032,523
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0479 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2073 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PANCTUT01
;      CLONE:  1515165
US-09-032-523-6

Alignment Scores:
Pred. No.:              39.2          Length:        2073
Score:                  56.00         Matches:       12
Percent Similarity:     62.07%        Conservative:  6
Best Local Similarity:  41.38%        Mismatches:    11
Query Match:            29.02%        Indels:        0
DB:                     4             Gaps:          0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-032-523-6 (1-2073)
```

```
Qy      2 GlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGlu 21
          |||   ||||||::::::|||           ||||||   |||:::::::|||   :::
Db    432 GGTCAATCTTTCTCCATTCAGTATGGAACCGGGAGCTTGTCCGGGATCATTGGAGCCGAC 491

Qy     22 TyrValLysLeuGluGlyLeuCysVal 30
            |||   :::   ||||||   |||
Db    492 CAAGTCTCTGTGGAAGGACTAACCGTG 518

RESULT 4
US-08-456-647B-19
; Sequence 19, Application US/08456647B
; Patent No. 5811516
;  GENERAL INFORMATION:
;    APPLICANT:  Lemke Ph.D. et al., Greg E.
;    TITLE OF INVENTION:  PROTEIN-TYROSINE KINASE GENES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,647B
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/237,401
;      FILING DATE:  02-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/884,486
;      FILING DATE:  15-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wetherell Ph.D., John R.
;      REGISTRATION NUMBER:  31,678
;      REFERENCE/DOCKET NUMBER:  07251/007002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3120 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    IMMEDIATE SOURCE:
;      CLONE:  Tyro-10
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  485..3047
US-08-456-647B-19

Alignment Scores:
Pred. No.:              66.6          Length:         3120
Score:                  56.00         Matches:        12
Percent Similarity:     54.84%        Conservative:   5
Best Local Similarity:  38.71%        Mismatches:     6
Query Match:            29.02%        Indels:         8
DB:                     1             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-08-456-647B-19 (1-3120)

Qy      3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
          ||||||:::|||||||||||||||                      ||||||
Db     91 GGGAGCTGGGCTCTTCAAGACTCA----------------------GGACCGAGGCAG 126

Qy     23 ValLysLeuLeuGlyLeuCysValCysLeuSer 33
          :::         |||:::   :::|||:::|||
Db    127 ATCTCATGTTTTGGGGTCTGGATTTGTGTCAGC 159

RESULT 5
US-08-237-401A-19
; Sequence 19, Application US/08237401A
; Patent No. 5837448
;  GENERAL INFORMATION:
;    APPLICANT:  Lemke Ph.D. et al., Greg E.
;    TITLE OF INVENTION:  PROTEIN-TYROSINE KINASE GENES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/237,401A
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/884,486
;      FILING DATE:  15-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07251/007001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3120 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    IMMEDIATE SOURCE:
;      CLONE:  Tyro-10
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  485..3047
US-08-237-401A-19

Alignment Scores:
Pred. No.:              66.6          Length:         3120
Score:                  56.00         Matches:        12
Percent Similarity:     54.84%        Conservative:   5
Best Local Similarity:  38.71%        Mismatches:     6
Query Match:            29.02%        Indels:         8
DB:                     2             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-08-237-401A-19 (1-3120)

Qy      3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
          ||||||:::|||||||||||||||                      ||||||
Db     91 GGGAGCTGGGCTCTTCAAGACTCA----------------------GGACCGAGGCAG 126

Qy     23 ValLysLeuLeuGlyLeuCysValCysLeuSer 33
          :::         |||:::   :::|||:::|||
Db    127 ATCTCATGTTTTGGGGTCTGGATTTGTGTCAGC 159

RESULT 6
US-08-875-944B-1/c
; Sequence 1, Application US/08875944B
```

```
; Patent No. 6096542
; GENERAL INFORMATION:
;   APPLICANT:  FUJINAGA, Kei
;   APPLICANT:  YOSHIDA, Koichi
;   APPLICANT:  HIGASHINO, Fumihiro
;   TITLE OF INVENTION:  CANCER CONTROL
;   NUMBER OF SEQUENCES:  5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;     STREET:  624 Ninth Street N.W., Ste. 300
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20001
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/875,944B
;     FILING DATE:  07-AUG-1997
;     CLASSIFICATION:  514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  JP 07-020173
;     FILING DATE:  08-FEB-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/JP96/00016
;     FILING DATE:  09-JAN-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BROWDY, Roger L.
;     REGISTRATION NUMBER:  25,618
;     REFERENCE/DOCKET NUMBER:  FUJINAGA=1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 628-5197
;     TELEFAX:  (202) 737-3528
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2064 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1..1386
US-08-875-944B-1

Alignment Scores:
Pred. No.:                66              Length:         2064
Score:                    54.50           Matches:        15
Percent Similarity:       44.44%          Conservative:   5
Best Local Similarity:    33.33%          Mismatches:     12
Query Match:              28.24%          Indels:         13
DB:                       3               Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-08-875-944B-1 (1-2064)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
           |||   |||   |||||||||         |||::::::||||||   |||
Db     278 GGCGGCTTCCTGCTGCAGGACAGGGCCGGGTCTGTGCGGGGACTCTGGGGCTCCTTCTTG 219

Qy      23 ValLysLeuLeuGlyLeu-------------------------------------Cys 29
           :::   ::::::||||||                                     |||
Db     218 ATCCTGGTGGTGGGGCTGTGGAAAGCTAGGTTTTCTGAATGGAAATCAGGAACAAACTGC 159

Qy      30 ValCysLeuSerGly 34
           |||||||||
Db     158 TCATCACTGTCTGGT 144

RESULT 7
US-09-116-049-3/c
; Sequence 3, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION:
;  APPLICANT: Hung, Men-Chie
;  TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
;  FILE REFERENCE: UTSC:582
;  CURRENT APPLICATION NUMBER: US/09/116,049A
;  CURRENT FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-116-049-3

Alignment Scores:
Pred. No.:                66              Length:         2064
Score:                    54.50           Matches:        15
Percent Similarity:       44.44%          Conservative:   5
Best Local Similarity:    33.33%          Mismatches:     12
Query Match:              28.24%          Indels:         13
DB:                       4               Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-116-049-3 (1-2064)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
           |||   |||   |||||||||         |||::::::||||||   |||
Db     278 GGCGGCTTCCTGCTGCAGGACAGGGCCGGGTCTGTGCGGGGACTCTGGGGCTCCTTCTTG 219

Qy      23 ValLysLeuLeuGlyLeu-------------------------------------Cys 29
           :::   ::::::||||||                                     |||
Db     218 ATCCTGGTGGTGGGGCTGTGGAAAGCTAGGTTTTCTGAATGGAAATCAGGAACAAACTGC 159

Qy      30 ValCysLeuSerGly 34
           |||||||||
Db     158 TCATCACTGTCTGGT 144

RESULT 8
US-08-780-835B-1/c
; Sequence 1, Application US/08780835B
; Patent No. 5922688
;  GENERAL INFORMATION:
;    APPLICANT:  Hung, Mien-Chie
;    APPLICANT:  Xing, Xiangming
;    TITLE OF INVENTION:  PEA3 is a Tumor Suppressor
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ARNOLD, WHITE AND DURKEE
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/780,835B
;      FILING DATE:  10-JAN-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  UTSC500
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2410 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-780-835B-1

Alignment Scores:
Pred. No.:              80.6          Length:         2410
Score:                  54.50         Matches:        15
Percent Similarity:     44.44%        Conservative:   5
Best Local Similarity:  33.33%        Mismatches:     12
Query Match:            28.24%        Indels:         13
DB:                     2             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-08-780-835B-1 (1-2410)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
           |||   |||   |||||||||         |||:::::::||||||   |||
Db     694 GGTGGCTTCCTGCTGCAGGACAGGGCGGGGTCTGTGCGGGGACTCTGGGGTTCCTTCTTG 635

Qy      23 ValLysLeuLeuGlyLeu-------------------------------------Cys 29
           :::   ::::::||||||                                      |||
Db     634 ATCCTGGTGGTGGGGCTATGGAAAGCTAAGTTTTCTGAATGGAAATCAGGAACAAACTGC 575

Qy      30 ValCysLeuSerGly 34
             |||||||||
Db     574 TCATCACTGTCCGGT 560

RESULT 9
US-09-303-268-1/c
; Sequence 1, Application US/09303268
; Patent No. 6172212
;   GENERAL INFORMATION:
;        APPLICANT: Hung, Mien-Chie
;                   Xing, Xiangming
;        TITLE OF INVENTION: PEA3 is a Tumor Suppressor
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: ARNOLD, WHITE AND DURKEE
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210-4433
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/303,268
;             FILING DATE: 30-Apr-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/780,835
;             FILING DATE: 10-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Wilson, Mark B.
;             REGISTRATION NUMBER: 37,259
;             REFERENCE/DOCKET NUMBER: UTSC500
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2410 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-303-268-1

Alignment Scores:
Pred. No.:              80.6          Length:         2410
Score:                  54.50         Matches:        15
Percent Similarity:     44.44%        Conservative:   5
Best Local Similarity:  33.33%        Mismatches:     12
Query Match:            28.24%        Indels:         13
DB:                     4             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-303-268-1 (1-2410)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
           |||   |||   |||||||||         |||:::::::||||||   |||
Db     694 GGTGGCTTCCTGCTGCAGGACAGGGCGGGGTCTGTGCGGGGACTCTGGGGTTCCTTCTTG 635

Qy      23 ValLysLeuLeuGlyLeu-------------------------------------Cys 29
           :::   ::::::||||||                                      |||
Db     634 ATCCTGGTGGTGGGGCTATGGAAAGCTAAGTTTTCTGAATGGAAATCAGGAACAAACTGC 575

Qy      30 ValCysLeuSerGly 34
             |||||||||
Db     574 TCATCACTGTCCGGT 560

RESULT 10
US-09-116-049-1/c
; Sequence 1, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION:
;  APPLICANT: Hung, Men-Chie
;  TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
;  FILE REFERENCE: UTSC:582
;  CURRENT APPLICATION NUMBER: US/09/116,049A
;  CURRENT FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2410
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-116-049-1

Alignment Scores:
Pred. No.:              80.6          Length:         2410
Score:                  54.50         Matches:        15
Percent Similarity:     44.44%        Conservative:   5
Best Local Similarity:  33.33%        Mismatches:     12
Query Match:            28.24%        Indels:         13
DB:                     4             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-116-049-1 (1-2410)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
           |||   |||   |||||||||         |||:::::::||||||   |||
Db     694 GGTGGCTTCCTGCTGCAGGACAGGGCGGGGTCTGTGCGGGGACTCTGGGGTTCCTTCTTG 635

Qy      23 ValLysLeuLeuGlyLeu-------------------------------------Cys 29
           :::   ::::::||||||                                      |||
Db     634 ATCCTGGTGGTGGGGCTATGGAAAGCTAAGTTTTCTGAATGGAAATCAGGAACAAACTGC 575

Qy      30 ValCysLeuSerGly 34
             |||||||||
Db     574 TCATCACTGTCCGGT 560

RESULT 11
US-09-676-610B-24/c
; Sequence 24, Application US/09676610B
; Patent No. 6444465
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT:  Jacqueline Wyatt
;  APPLICANT:  Susan M. Freier
;  TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
;  FILE REFERENCE: RTS-0138
;  CURRENT APPLICATION NUMBER: US/09/676,610B
;  CURRENT FILING DATE:  2000-09-29
;  NUMBER OF SEQ ID NOS: 182
; SEQ ID NO 24
```

```
;   LENGTH: 169998
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1208)...(1472)
;   NAME/KEY: intron
;   LOCATION: (1473)...(124390)
;   NAME/KEY: exon
;   LOCATION: (124391)...(124544)
;   NAME/KEY: intron
;   LOCATION: (124545)...(125409)
;   NAME/KEY: exon
;   LOCATION: (125410)...(125595)
;   NAME/KEY: intron
;   LOCATION: (125596)...(128711)
;   NAME/KEY: exon
;   LOCATION: (128712)...(128848)
    NAME/KEY: intron
    LOCATION: (128849)...(133400)
    NAME/KEY: exon
;   LOCATION: (133401)...(133469)
;   NAME/KEY: intron
;   LOCATION: (133470)...(134652)
;   NAME/KEY: exon
;   LOCATION: (134653)...(134773)
;   NAME/KEY: intron
;   LOCATION: (134774)...(136116)
;   NAME/KEY: exon
;   LOCATION: (136117)...(136261)
;   NAME/KEY: intron
;   LOCATION: (136262)...(137936)
;   NAME/KEY: exon
;   LOCATION: (137937)...(138053)
;   NAME/KEY: intron
;   LOCATION: (138054)...(138637)
;   NAME/KEY: exon
;   LOCATION: (138638)...(138766)
;   NAME/KEY: intron
;   LOCATION: (138767)...(138864)
;   NAME/KEY: exon
;   LOCATION: (138865)...(138940)
;   NAME/KEY: intron
;   LOCATION: (138941)...(139765)
;   NAME/KEY: exon
;   LOCATION: (139766)...(139860)
;   NAME/KEY: intron
    LOCATION: (139861)...(142245)
    NAME/KEY: exon
,   LOCATION: (142246)...(142445)
;   NAME/KEY: intron
;   LOCATION: (142446)...(143605)
;   NAME/KEY: exon
;   LOCATION: (143606)...(143738)
;   NAME/KEY: intron
;   LOCATION: (143739)...(145838)
;   NAME/KEY: exon
;   LOCATION: (145839)...(145931)
;   NAME/KEY: intron
;   LOCATION: (145932)...(147385)
;   NAME/KEY: exon
;   LOCATION: (147386)...(147544)
;   NAME/KEY: intron
;   LOCATION: (147545)...(153274)
;   NAME/KEY: exon
;   LOCATION: (153275)...(153321)
;   NAME/KEY: intron
;   LOCATION: (153322)...(155088)
;   NAME/KEY: exon
;   LOCATION: (155089)...(155231)
;   NAME/KEY: intron
;   LOCATION: (155232)...(156025)
;   NAME/KEY: exon
;   LOCATION: (156026)...(156151)
;   NAME/KEY: intron
;   LOCATION: (156152)...(156826)
;   NAME/KEY: exon
;   LOCATION: (156827)...(156928)
;   NAME/KEY: intron
;   LOCATION: (156929)...(163399)
;   NAME/KEY: exon
;   LOCATION: (163400)...(163586)
US-09-676-610B-24

Alignment Scores:
Pred. No.:              2.32e+04        Length:         169998
Score:                  54.00           Matches:        12
Percent Similarity:     51.35%          Conservative:   7
Best Local Similarity:  32.43%          Mismatches:     14
Query Match:            27.98%          Indels:         4
DB:                     4               Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-676-610B-24 (1-169998)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGly--------- 19
           |||      :::      |||      ::::::   :::||||||:::|||
Db   98748 GGTATGGTCAGCTGTCCTGATCTGGGTGCCACTTACAAAGGGCTAATTGGTTTGTGATAT 98689

Qy      20 ---ProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
             |||   |||      |||      :::|||||||||:::      |||
Db   98688 TGGCCAAATTACACACCTTTGACTTGTGTGTGTGTGTGTGTGTATGTGT 98638

RESULT 12
US-09-526-993-4/c
; Sequence 4, Application US/09526993
; Patent No. 6465715
; GENERAL INFORMATION:
;  APPLICANT: Zwaal, Richard
;  APPLICANT: Asaert, Wouter
;  APPLICANT: Roelens, Ingele
;  APPLICANT: Bogaert, Thierry
;  TITLE OF INVENTION: EXPRESSION OF DNA OR PROTEINS IN C. ELEGANS
;  FILE REFERENCE: B0192/7012/ERG/KA
;  CURRENT APPLICATION NUMBER: US/09/526,993
;  CURRENT FILING DATE: 2000-03-16
;  EARLIER APPLICATION NUMBER: U.K. 9906018.8
;  EARLIER FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 1308
;   TYPE: DNA
;   ORGANISM: Caenorhabditis Elegans
US-09-526-993-4

Alignment Scores:
Pred. No.:              52              Length:         1308
Score:                  53.50           Matches:        11
Percent Similarity:     64.29%          Conservative:   7
Best Local Similarity:  39.29%          Mismatches:     7
Query Match:            27.72%          Indels:         3
DB:                     4               Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-526-993-4 (1-1308)

Qy       5 PheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLys 24
           |||      |||:::||||||   |||:::::::           |||:::   ::::::
Db     194 TTTGAATTTCAAAACAGTTTCGTTTCAATCGAA---------ATTCCCGATATCATTCGC 144

Qy      25 LeuLeuGlyLeuCysValCysLeu 32
           ||||||   :::|||   ||||||
Db     143 CTCTTGAATATTTGTCGATGTTTG 120

RESULT 13
US-09-526-993-3/c
; Sequence 3, Application US/09526993
```

```
; Patent No. 6465715
; GENERAL INFORMATION:
;  APPLICANT: Zwaal, Richard
;  APPLICANT: Asaert, Wouter
;  APPLICANT: Roelens, Ingele
;  APPLICANT: Bogaert, Thierry
;  TITLE OF INVENTION: EXPRESSION OF DNA OR PROTEINS IN C. ELEGANS
;  FILE REFERENCE: B0192/7012/ERG/KA
;  CURRENT APPLICATION NUMBER: US/09/526,993
;  CURRENT FILING DATE: 2000-03-16
;  EARLIER APPLICATION NUMBER: U.K. 9906018.8
;  EARLIER FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2167
;   TYPE: DNA
;   ORGANISM: Caenorhabditis Elegans
US-09-526-993-3

Alignment Scores:
Pred. No.:                99.8          Length:        2167
Score:                    53.50         Matches:       11
Percent Similarity:       64.29%        Conservative:  7
Best Local Similarity:    39.29%        Mismatches:    7
Query Match:              27.72%        Indels:        3
DB:                       4             Gaps:          1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-526-993-3 (1-2167)

Qy       5 PheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLys 24
           |||      |||:::||||||   |||::::::          ||:::   ::::::
Db     194 TTTGAATTTCAAAACAGTTTCGTTTCAATCGAA---------ATTCCCGATATCATTCGC 144

Qy      25 LeuLeuGlyLeuCysValCysLeu 32
           ||||||   :::|||   ||||||
Db     143 CTCTTGAATATTTGTCGATGTTTG 120

RESULT 14
US-07-778-890A-2/c
; Sequence 2, Application US/07778890A
; Patent No. 5266489
;  GENERAL INFORMATION:
;    APPLICANT:  1.REY-SENELONGE Arielle
;    APPLICANT:  2.KOHEN Gilla
;    TITLE OF INVENTION:  recombinant herpes viruses,in particular for the
;    TITLE OF INVENTION:  production of vaccines,process for preparing them,plasmids prc
     TITLE OF INVENTION:  during this process and vaccines obtained.
     NUMBER OF SEQUENCES:  7
     CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Larson and Taylor
;      STREET:  727 Twenty-Third Street,South
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-5.25 inch,360Kb storage
;      COMPUTER:  IBM PC/XT/AT or compatibles
;      OPERATING SYSTEM:  PC-DOS or MS-DOS version 2.2 or above
;      SOFTWARE:  KEDIT or any ASCII Text Editor
;    CURRENT APPLICATION DATA:  07/778,890
;      APPLICATION NUMBER:  US/07/778,890A
;      FILING DATE:  19920103
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR9003105,PCT/FR/91/00184
;      FILING DATE:  12-MAR-1990,07-MAR-1991
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2176 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  cDNA to genomic RNA
;    HYPOTHETICAL:  yes
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  Newcastle disease virus(NDV)
;      STRAIN:  Texas strain
;      INDIVIDUAL ISOLATE:  chicken
;    IMMEDIATE SOURCE:
;      LIBRARY:  plasmid
;      CLONE:  pNDV 108
;    FEATURE:
;      NAME/KEY:  part of matrix protein gene and fusion protein gene
;      LOCATION:  from     to      description
;      LOCATION:  1        271     matrix protein gene
;      LOCATION:  431      2092    fusion protein gene
US-07-778-890A-2

Alignment Scores:
Pred. No.:                100           Length:        2176
Score:                    53.50         Matches:       14
Percent Similarity:       55.88%        Conservative:  5
Best Local Similarity:    41.18%        Mismatches:    12
Query Match:              27.72%        Indels:        3
DB:                       1             Gaps:          2

US-09-513-999C-7869_COPY_1_37 (1-37) x US-07-778-890A-2 (1-2176)

Qy       7 LeuGlnAspSerPheSerSerLeuGlnGly---LeuLeuGlyProGluTyrValLysLeu 25
           |||   |||   |||   :::   |||         |||   ||||||||||:::
Db     166 TTGAATGATGATTTTTACACTCCGCAGGCGCGCAGTTTGACTCCAGAGTATCTTAGCTAC 107

Qy      26 Leu------GlyLeuCysValCysLeuSerGlyCysSerThr 37
           |||      |||:::|||      ::::::||||||   |||
Db     106 CTGAGGAGAGGCATTTGCTATAGGATAGCAGGCTGTCCCACT 65

RESULT 15
US-09-526-993-2/c
; Sequence 2, Application US/09526993
; Patent No. 6465715
; GENERAL INFORMATION:
;  APPLICANT: Zwaal, Richard
;  APPLICANT: Asaert, Wouter
;  APPLICANT: Roelens, Ingele
;  APPLICANT: Bogaert, Thierry
;  TITLE OF INVENTION: EXPRESSION OF DNA OR PROTEINS IN C. ELEGANS
;  FILE REFERENCE: B0192/7012/ERG/KA
;  CURRENT APPLICATION NUMBER: US/09/526,993
;  CURRENT FILING DATE: 2000-03-16
;  EARLIER APPLICATION NUMBER: U.K. 9906018.8
;  EARLIER FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 3224
;   TYPE: DNA
;   ORGANISM: Caenorhabditis Elegans
US-09-526-993-2

Alignment Scores:
Pred. No.:                167           Length:        3224
Score:                    53.50         Matches:       11
Percent Similarity:       64.29%        Conservative:  7
Best Local Similarity:    39.29%        Mismatches:    7
Query Match:              27.72%        Indels:        3
DB:                       4             Gaps:          1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-526-993-2 (1-3224)

Qy       5 PheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLys 24
           |||      |||:::||||||   |||::::::          ||:::   ::::::
Db    1251 TTTGAATTTCAAAACAGTTTCGTTTCAATCGAA---------ATTCCCGATATCATTCGC 1201
```

```
Qy      25 LeuLeuGlyLeuCysValCysLeu 32
           ||||||   :::|||   ||||||
Db    1200 CTCTTGAATATTTGTCGATGTTTG 1177
```

Search completed: April 24, 2003, 23:20:43
Job time : 58 secs

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2003, 23:01:58 ; Search time 77 Seconds
(without alignments)
522.867 Million cell updates/sec

Title: US-09-513-999C-7869_COPY_1_37
Perfect score: 193
Sequence: 1 MGGSFALQDSFSSLQGLLGPEYVKLLGLCVCLSGCST 37

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09513999/runat_18042003_170939_28423/app_query.fasta_1.199
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09513999_@CGN_1_1_77_@runat_18042003_170939_28423
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	93	48.2	684973	10	US-09-263-959-1	Sequence 1, Appli
c	2	88.5	45.9	570	10	US-09-864-761-9118	Sequence 9118, Ap
	3	78	40.4	428	9	US-09-918-995-8153	Sequence 8153, Ap
	4	69.5	36.0	412	10	US-09-783-590-1450	Sequence 1450, Ap
	5	63	32.6	162	10	US-09-864-761-31393	Sequence 31393, A
	6	63	32.6	519	10	US-09-864-761-14866	Sequence 14866, A
	7	60.5	31.3	2730	9	US-09-965-528-27	Sequence 27, Appl
	8	60.5	31.3	24533	9	US-09-764-868-1349	Sequence 1349, Ap
	9	59.5	30.8	506	10	US-09-783-590-5991	Sequence 5991, Ap
c	10	57.5	29.8	1605	10	US-09-930-218-10	Sequence 10, Appl
	11	57	29.5	1155	10	US-09-833-381-8	Sequence 8, Appli
	12	56	29.0	1380	10	US-09-974-300-685	Sequence 685, App
	13	56	29.0	3120	9	US-09-158-722-19	Sequence 19, Appl
	14	56	29.0	302250	10	US-09-962-832-154	Sequence 154, App
c	15	55	28.5	250	10	US-09-044-604-11	Sequence 11, Appl
c	16	55	28.5	360	10	US-09-867-701-8266	Sequence 8266, Ap
c	17	55	28.5	435	9	US-09-796-692-4049	Sequence 4049, Ap
	18	55	28.5	467	10	US-09-864-761-6467	Sequence 6467, Ap
	19	55	28.5	747	12	US-10-001-879-101	Sequence 101, App
c	20	55	28.5	855	10	US-09-044-604-1	Sequence 1, Appli
	21	55	28.5	3824	9	US-10-036-041-22	Sequence 22, Appl
	22	55	28.5	3824	9	US-10-028-072-541	Sequence 541, App
	23	55	28.5	3824	9	US-10-035-855-22	Sequence 22, Appl
	24	55	28.5	3824	9	US-10-121-049-541	Sequence 541, App
	25	55	28.5	3824	9	US-10-123-904-541	Sequence 541, App
	26	55	28.5	3824	9	US-10-140-470-541	Sequence 541, App
	27	55	28.5	3824	9	US-09-931-836-22	Sequence 22, Appl
	28	55	28.5	3824	9	US-10-175-746-541	Sequence 541, App
	29	55	28.5	3824	9	US-10-176-918-541	Sequence 541, App
	30	55	28.5	3824	9	US-10-176-921-541	Sequence 541, App
	31	55	28.5	3824	9	US-10-227-884-209	Sequence 209, App
	32	55	28.5	3824	9	US-10-036-214-22	Sequence 22, Appl
	33	55	28.5	3824	9	US-10-137-865-541	Sequence 541, App
	34	55	28.5	3824	9	US-10-140-474-541	Sequence 541, App
	35	55	28.5	3824	9	US-10-035-719-22	Sequence 22, Appl
	36	55	28.5	3824	9	US-10-142-431-541	Sequence 541, App
	37	55	28.5	3824	9	US-10-143-114-541	Sequence 541, App
	38	55	28.5	3824	9	US-10-230-163-209	Sequence 209, App
	39	55	28.5	3824	9	US-10-140-002-541	Sequence 541, App
	40	55	28.5	3824	9	US-10-036-160-22	Sequence 22, Appl
	41	55	28.5	3824	9	US-10-142-419-541	Sequence 541, App
	42	55	28.5	3824	9	US-10-218-631-209	Sequence 209, App
	43	55	28.5	3824	9	US-10-230-338-209	Sequence 209, App
	44	55	28.5	3824	9	US-10-035-958-22	Sequence 22, Appl
	45	55	28.5	3824	9	US-10-036-150-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.

```
;       REGISTRATION NUMBER:  33,963
;       REFERENCE/DOCKET NUMBER:  920010.426C2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  684973 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-263-959-1

Alignment Scores:
Pred. No.:              0.0381          Length:         684973
Score:                  93.00           Matches:        19
Percent Similarity:     75.00%          Conservative:   2
Best Local Similarity:  67.86%          Mismatches:     7
Query Match:            48.19%          Indels:         0
                        10              Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-263-959-1 (1-684973)

Qy       9 AspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeu 28
           |||   |||   :::|||   |||:::|||     |||  ||||||||||||||||||
Db 404731 GACAGATTTCCTGCCTTGCTGGGGATCCTGAGGCTGGAATGTGTAAAACTCCTGGGTCTC 404672

Qy      29 CysValCysLeuSerGlyCysSer 36
           ||||||||||||||  ||||||
Db 404671 TGTGTGTGCCTGAGCAGCTGCTCT 404648

RESULT 2
US-09-864-761-9118/c
; Sequence 9118, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
   PRIOR APPLICATION NUMBER: US 60/207,456
   PRIOR FILING DATE: 2000-05-26
   PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9118
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000053.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
US-09-864-761-9118

Alignment Scores:
Pred. No.:              2.61e-05        Length:         570
Score:                  88.50           Matches:        21
Percent Similarity:     61.11%          Conservative:   1
Best Local Similarity:  58.33%          Mismatches:     7
Query Match:            45.85%          Indels:         7
DB:                     10              Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-864-761-9118 (1-570)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           |||   ||||||                          :::|||   |||   ||||||
Db     568 ATGAATGGATCT--------------------CCTGCCTTGCTGGGATTCCTTGGGCTG 530

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCysSer 36
           ||||||||||||   ||||||||||||||||||||||||||||||   |||
Db     529 GAGTATGTAAAATTCCTGGGTCTTTGTGTGTGCCTGAGTGGCCGCTCT 482

RESULT 3
US-09-918-995-8153
; Sequence 8153, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8153
;   LENGTH: 428
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-8153

Alignment Scores:
Pred. No.:              0.000899        Length:         428
Score:                  78.00           Matches:        17
Percent Similarity:     71.43%          Conservative:   3
Best Local Similarity:  60.71%          Mismatches:     8
Query Match:            40.41%          Indels:         0
DB:                     9               Gaps:           0
```

```
US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-918-995-8153 (1-428)

Qy      6 AlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyrValLysLeu 25
          ||||||   :::      |||         |||:::|||     ||||||||||||||||
Db    273 GCCTTGCCTGAGATGCCATCACCTTTCTTGGGGATCCTAAGGCTGGAGTATGTAAAGCTC 332

Qy     26 LeuGlyLeuCysValCysLeuSer 33
          ||||||||||||:::|||||||||
Db    333 CTGGGTCTCTGTATGTGCCTGAGC 356

RESULT 4
US-09-783-590-1450
; Sequence 1450, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1450
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (5)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (189)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (190)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (231)
;   OTHER INFORMATION: n equals a,t,g, or c
    NAME/KEY: misc feature
    LOCATION: (265)
    OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (266)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (268)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (269)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (274)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (275)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (339)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (340)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (370)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (399)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (402)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (403)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (406)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1450

Alignment Scores:
Pred. No.:              0.0202          Length:         412
Score:                  69.50           Matches:        16
Percent Similarity:     61.29%          Conservative:   3
Best Local Similarity:  51.61%          Mismatches:     7
Query Match:            36.01%          Indels:         5
DB:                     10              Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-783-590-1450 (1-412)

Qy     11 PheSerSerLeuGln----------GlyLeuLeuGlyProGluTyrValLysLeu 25
          |||:::   ||||||               |||     |||||||||:::::|||
Db    172 TTTGCTGAGTTGCAGCTNNTTTTTCTGGGAAGCCTGGAAAAGCCAGAGTATCTGAGGCTN 231

Qy     26 LeuGlyLeuCysValCysLeuSerGlyCysSer 36
          |||   ||||||     ||||||||||||||||
Db    232 TTGAATCTGTGCACAGGCCTCAGTGGCTGCTCT 264

RESULT 5
US-09-864-761-31393
; Sequence 31393, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
```

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31393
;   LENGTH: 162
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC017089.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
    OTHER INFORMATION: SWISSPROT HIT: P98161, EVALUE 3.00e+00
    OTHER INFORMATION: EST_HUMAN HIT: AI792950.1, EVALUE 6.00e-07
;   OTHER INFORMATION: NT HIT: AL163210.2, EVALUE 4.00e-04
US-09-864-761-31393

Alignment Scores:
Pred. No.:                0.0695          Length:          162
Score:                    63.00           Matches:         13
Percent Similarity:       66.67%          Conservative:    1
Best Local Similarity:    61.90%          Mismatches:      7
Query Match:              32.64%          Indels:          0
DB:                       10              Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-864-761-31393 (1-162)

Qy      16 GlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
           |||      |||||||||||||:::|||||||          ||||||||||||   |||
Db      60 GGAAGTCCAGGGCCAGAATATCTAAAGCTCTGGAGTTTTCATGTGTGCCTGAGCAGATGC 119

Qy      36 Ser 36
           |||
Db     120 TCT 122

RESULT 6
US-09-864-761-14866
; Sequence 14866, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14866
;   LENGTH: 519
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC017089.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
US-09-864-761-14866

Alignment Scores:
Pred. No.:                0.303           Length:          519
Score:                    63.00           Matches:         13
Percent Similarity:       66.67%          Conservative:    1
Best Local Similarity:    61.90%          Mismatches:      7
Query Match:              32.64%          Indels:          0
DB:                       10              Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-864-761-14866 (1-519)

Qy      16 GlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
           |||      |||||||||||||:::|||||||          ||||||||||||   |||
Db     309 GGAAGTCCAGGGCCAGAATATCTAAAGCTCTGGAGTTTTCATGTGTGCCTGAGCAGATGC 368

Qy      36 Ser 36
           |||
Db     369 TCT 371

RESULT 7
US-09-965-528-27
; Sequence 27, Application US/09965528
; Publication No. US20020187523A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  YUE, Henry
;  APPLICANT:  LAL, Preeti
;  APPLICANT:  BURFORD, Neil
;  APPLICANT:  BANDMAN, Olga
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra
;  TITLE OF INVENTION: EXTRACELLULAR SIGNALING MOLECULES
;  FILE REFERENCE: PF-0701 USA
;  CURRENT APPLICATION NUMBER: US/09/965,528
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: 60/134,949
;  PRIOR FILING DATE: 1999-05-19
;  PRIOR APPLICATION NUMBER: 60/144,270
;  PRIOR FILING DATE: 1999-07-15
```

```
;  PRIOR APPLICATION NUMBER: 60/146,700
;  PRIOR FILING DATE: 1999-07-30
;  PRIOR APPLICATION NUMBER: 60/157,508
;  PRIOR FILING DATE: 1999-10-04
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PERL Program
; SEQ ID NO 27
;   LENGTH: 2730
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020187523A1 1288847CB1
US-09-965-528-27

Alignment Scores:
Pred. No.:                6.24          Length:        2730
Score:                    60.50         Matches:       13
Percent Similarity:       59.26%        Conservative:  3
Best Local Similarity:    48.15%        Mismatches:    10
Query Match:              31.35%        Indels:        1
DB:                       9             Gaps:          1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-965-528-27 (1-2730)

Qy      10 SerPheSerSerLeuGlnGlyLeuLeu---GlyProGluTyrValLysLeuLeuGlyLeu 28
           :::|||         ||||||||||||      |||   |||   |||||||:::|||
Db     630 ACCTTCCTAGGCTCTCAGGGGCTCCTTCAGTCCCCATCCTATGAGAAACTAGTGGGTTGC 689

Qy      29 CysValCysLeuSerGlyCys 35
           |||:::       ||||||
Db     690 TGCCTGATGACAAGGGGTTGT 710

RESULT 8
US-09-764-868-1349
; Sequence 1349, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ32
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1349
    LENGTH: 24533
    TYPE: DNA
    ORGANISM: Homo sapiens
US-09-764-868-1349

Alignment Scores:
Pred. No.:                100           Length:        24533
Score:                    60.50         Matches:       16
Percent Similarity:       60.00%        Conservative:  5
Best Local Similarity:    45.71%        Mismatches:    13
Query Match:              31.35%        Indels:        1
DB:                       9             Gaps:          1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-764-868-1349 (1-24533)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyPro 20
           :::|||   :::   ||||||     :::          |||   ||||||||||||
Db   14224 CTGGGCTGGGCTGTGGCCCTTGCTAGCGCTCGTGGTGGCGCCCAGGAGCTTTTGGGTCCT 14283

Qy      21 GluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
           |||   ::::::|||||||||||||||            ||||||
Db   14284 GAG---ATGCAACTGCTTGGACTGTGCCGGGGATAGATAGGCTGC 14325

RESULT 9
US-09-783-590-5991
; Sequence 5991, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5991
;   LENGTH: 506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (110)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (331)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (351)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (364)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (402)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (435)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (460)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (502)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5991

Alignment Scores:
Pred. No.:                1.08          Length:        506
Score:                    59.50         Matches:       16
Percent Similarity:       55.26%        Conservative:  5
Best Local Similarity:    42.11%        Mismatches:    8
Query Match:              30.83%        Indels:        9
DB:                       10            Gaps:          2

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-783-590-5991 (1-506)

Qy       3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeu---------- 18
           |||         :::::::|||   ||||||      :::|||||||||
Db     134 GGAGAAAAAGGAGTCAAGGATTAGTTCAGCTTCTGCCGTGGCCTTCTCAGCAAGTGGTGG 193

Qy      19 ----------GlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeu 32
                     |||   ||||||   |||:::||||||   |||:::||||||
Db     194 GGAATTCACTGAGGTGAGGAATAC---AAAATACTTGGAGGTTGCATCTGTTTG 244

RESULT 10
US-09-930-218-10/c
; Sequence 10, Application US/09930218
; Patent No. US20020034810A1
; GENERAL INFORMATION:
;  APPLICANT: goldshmidt, orit
;  APPLICANT:  pecker, iris
```

```
;  APPLICANT:  vlodavsky, israel
;  APPLICANT:  israel, michal
;  TITLE OF INVENTION: AVIAN AND REPTILE DERIVED POLYNUCLEOTIDE ENCODING A POLYPEPTIDE H
;  TITLE OF INVENTION:  HEPARANASE ACTIVITY
;  FILE REFERENCE: 01/22335
;  CURRENT APPLICATION NUMBER: US/09/930,218
;  CURRENT FILING DATE:  2001-08-16
;  PRIOR APPLICATION NUMBER: 09/666,390
;  PRIOR FILING DATE: 2000-09-20
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 1605
;   TYPE: DNA
;   ORGANISM: Gallus gallus
US-09-930-218-10

Alignment Scores:
Pred. No.:                9.73          Length:          1605
Score:                    57.50         Matches:         11
Percent Similarity:       73.68%        Conservative:    3
Best Local Similarity:    57.89%        Mismatches:      4
Query Match:              29.79%        Indels:          1
DB:                       10            Gaps:            1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-930-218-10 (1-1605)

Qy      20 ProGluTyrValLysLeuLeuGlyLeuCys---ValCysLeuSerGlyCysSerThr 37
           |||:::::::::|||||||||      |||   |||||||||   ||||||   |||
Db     819 CCAGATTTCATGAAGCTTCTGAGCAGGTGCTGGGTGTGCTTGCGGGGCTGCCCCACG 763

RESULT 11
US-09-833-381-8
; Sequence 8, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
;  FILE REFERENCE: 5800-119
;  CURRENT APPLICATION NUMBER: US/09/833,381
;  CURRENT FILING DATE:  2001-04-11
;  PRIOR APPLICATION NUMBER: 09/516,448
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 2050
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-381-8

Alignment Scores:
Pred. No.:                7.73          Length:          1155
Score:                    57.00         Matches:         12
Percent Similarity:       53.66%        Conservative:    10
Best Local Similarity:    29.27%        Mismatches:      13
Query Match:              29.53%        Indels:          6
DB:                       10            Gaps:            1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-833-381-8 (1-1155)

Qy       1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeu----------------- 14
           :::||||||::::::   |||      :::   :::::::|||
Db     535 GTAGGAGGCACCTATACCCTTGTACATGCCAAAAATGCCCTCGGTCCGAGCTGTCTGCAG 594

Qy      15 GlnGlyLeuLeuGlyProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGly 34
           |||   :::      |||         :::   |||   ||||||:::|||   :::|||
Db     595 CAGAGCGTACAGTATCCCTCGGTACATGAGGCCCTTGCCCTGTGCATCTGTGCGGCTGGT 654

Qy      35 Cys 35
           |||
Db     655 TGT 657

RESULT 12
US-09-974-300-685
; Sequence 685, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 685
;   LENGTH: 1380
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-685

Alignment Scores:
Pred. No.:                14            Length:          1380
Score:                    56.00         Matches:         9
Percent Similarity:       75.00%        Conservative:    3
Best Local Similarity:    56.25%        Mismatches:      4
Query Match:              29.02%        Indels:          0
DB:                       10            Gaps:            0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-974-300-685 (1-1380)

Qy      20 ProGluTyrValLysLeuLeuGlyLeuCysValCysLeuSerGlyCys 35
           ||||||   |||:::   ||||||   |||   |||::::::||||||
Db     564 CCCGAGCTTGTTCAATCCTTGGGGACTTGCAACTGCATCACCGGATGT 611

RESULT 13
US-09-158-722-19
; Sequence 19, Application US/09158722
; Publication No. US20030013848A1
;  GENERAL INFORMATION:
;    APPLICANT:  Lemke Ph.D. et al., Greg E.
;    TITLE OF INVENTION:  PROTEIN-TYROSINE KINASE GENES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/158,722
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/456,647
;      FILING DATE:  02-JUN-1995
;      APPLICATION NUMBER:  US 08/237,401
;      FILING DATE:  02-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/884,486
;      FILING DATE:  15-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wetherell Ph.D., John R.
;      REGISTRATION NUMBER:  31,678
```

```
;      REFERENCE/DOCKET NUMBER:  07251/007002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3120 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    IMMEDIATE SOURCE:
;      CLONE:  Tyro-10
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  485..3047
US-09-158-722-19

Alignment Scores:
Pred. No.:              39.4          Length:         3120
Score:                  56.00         Matches:        12
Percent Similarity:     54.84%        Conservative:   5
Best Local Similarity:  38.71%        Mismatches:     6
Query Match:            29.02%        Indels:         8
DB:                     9             Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-158-722-19 (1-3120)

Qy      3 GlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGluTyr 22
          ||||||:::||||||||||||||||                      ||||||
Db     91 GGGAGCTGGGCTCTTCAAGACTCA----------------------GGACCGAGGCAG 126

Qy     23 ValLysLeuLeuGlyLeuCysValCysLeuSer 33
          :::       |||:::  :::|||:::|||
Db    127 ATCTCATGTTTTGGGGTCTGGATTTGTGTCAGC 159

RESULT 14
US-09-962-832-154
; Sequence 154, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
;  APPLICANT: Ebner, Reinhard
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-74
;  CURRENT APPLICATION NUMBER: US/09/962,832
;  CURRENT FILING DATE:  2001-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,077
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,280
;  PRIOR FILING DATE: 2000-09-25
;  NUMBER OF SEQ ID NOS: 259
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 154
;   LENGTH: 302250
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-962-832-154

Alignment Scores:
Pred. No.:              1.27e+04      Length:         302250
Score:                  56.00         Matches:        12
Percent Similarity:     61.29%        Conservative:   7
Best Local Similarity:  38.71%        Mismatches:     10
Query Match:            29.02%        Indels:         2
DB:                     10            Gaps:           1

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-962-832-154 (1-302250)

Qy      7 LeuGlnAspSerPheSerSerLeuGlnGlyLeu-----LeuGlyProGluTyrValLys 24
          |||:::::|||   :::||||||      :::      ||||||
Db 174135 CTTGAGGAGTCTGTGAACTCCCTAAATTCGATGCAGCCTCTGGGGGTGTGGGTGTGTGTG 174194

Qy     25 LeuLeuGlyLeuCysValCysLeuSerGlyCys 35
          :::|||:::||||||||||:::   ||||||
Db 174195 TGTGTGGGGGTGTGTGTGTGTTTGTGGGTGT 174227

RESULT 15
US-09-044-604-11/c
; Sequence 11, Application US/09044604
; Patent No. US20020009718A1
;  GENERAL INFORMATION:
;    APPLICANT:  Timms, Kathy L.
;    TITLE OF INVENTION:  ENDOMETRIOSIS-SPECIFIC SECRETORY
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KOHN & ASSOCIATES
;      STREET:  30500 No. US20020009718A1thwestern Hwy.  Suite 410
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/044,604
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Montgomery, Ilene N.
;      REGISTRATION NUMBER:  38,972
;      REFERENCE/DOCKET NUMBER:  0994.00084
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (248) 539-5050
;      TELEFAX:  (248) 539-5055
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  250 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Probefor human ENDO-I"
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-044-604-11

Alignment Scores:
Pred. No.:              2.35          Length:         250
Score:                  55.00         Matches:        12
Percent Similarity:     51.61%        Conservative:   4
Best Local Similarity:  38.71%        Mismatches:     15
Query Match:            28.50%        Indels:         0
DB:                     10            Gaps:           0

US-09-513-999C-7869_COPY_1_37 (1-37) x US-09-044-604-11 (1-250)

Qy      2 GlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuLeuGlyProGlu 21
          |||   |||:::   ::::::      |||   ||||||||||||         |||
Db    191 GGGAGTAGTTGGGGTGTAGAACAACCTTTTCAATCTCTACAAGGTGCTTTTTCCCCACAT 132

Qy     22 TyrValLysLeuLeuGlyLeuCysValCysLeu 32
             |||          |||   |||:::   |||
Db    131 AGAGTGTTAAAGTAGGGGCAATGTCTTTCGCTG 99


Search completed: April 24, 2003, 23:57:24
Job time : 153 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 00:54:24 ; Search time 938 Seconds
(without alignments)
3443.935 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

al number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	111	100.0	169620	2	AC012674	AC012674 Homo sapi
c	2	66.8	60.2	199038	2	AC116565	AC116565 Homo sapi
	3	65.2	58.7	130636	9	AC092586	AC092586 Homo sapi
c	4	64.6	58.2	140040	9	AC022029	AC022029 Homo sapi
	5	64.6	58.2	147288	9	AL355812	AL355812 Human DNA
c	6	63.6	57.3	75270	9	AL591048	AL591048 Human DNA
c	7	63.6	57.3	143372	9	AL137847	AL137847 Human DNA
c	8	63.6	57.3	153053	9	AL137179	AL137179 Human DNA
	9	63.6	57.3	153422	2	AC023408	AC023408 Homo sapi
c	10	63.4	57.1	86314	9	AC109592	AC109592 Homo sapi
c	11	63.4	57.1	159475	2	AC021378	AC021378 Homo sapi
c	12	63	56.8	153783	9	AL157937	AL157937 Human DNA
c	13	62.4	56.2	123779	30	AC021025	Ac021025 Homo sapi
	14	62.4	56.2	128118	2	AC076969	AC076969 Homo sapi
c	15	62.4	56.2	128583	9	AC121249	AC121249 Homo sapi
	16	62.4	56.2	131347	2	AC002421	AC002421 Homo sapi
	17	62.4	56.2	170623	9	AL391375	AL391375 Human DNA
	18	62.4	56.2	178650	9	AC104303	AC104303 Homo sapi
	19	61.4	55.3	21104	9	AC110811	AC110811 Homo sapi
	20	61.4	55.3	121902	9	AC008386	AC008386 Homo sapi
	21	61.4	55.3	163083	9	AC103773	AC103773 Homo sapi
c	22	61.4	55.3	163231	9	AL162575	AL162575 Human DNA
c	23	61.4	55.3	164819	9	AC092335	AC092335 Homo sapi
c	24	61.4	55.3	170736	2	AC013658	AC013658 Homo sapi
c	25	61.4	55.3	177423	9	AC120036	AC120036 Homo sapi
	26	61.4	55.3	184886	9	AC016650	AC016650 Homo sapi
	27	61.4	55.3	187847	2	AC023155	AC023155 Homo sapi
c	28	61.4	55.3	196145	9	AC026241	AC026241 Homo sapi
	29	60.8	54.8	38936	9	AL358817	AL358817 Human DNA
c	30	60.8	54.8	169772	9	AC069538	AC069538 Homo sapi
	31	60.4	54.4	32918	2	AC007445	AC007445 Homo sapi
	32	60.4	54.4	68137	2	AC079966	AC079966 Homo sapi
c	33	60.4	54.4	159747	2	AP001019	AP001019 Homo sapi
c	34	60.4	54.4	163065	9	AC021483	AC021483 Homo sapi
c	35	60.4	54.4	168062	30	AC022292	Ac022292 Homo sapi
c	36	60.4	54.4	169462	9	AC026826	AC026826 Homo sapi
	37	60.4	54.4	186224	9	AC105901	AC105901 Homo sapi
c	38	60.4	54.4	200333	2	AC073279	AC073279 Homo sapi
c	39	60	54.1	148290	9	AC107979	AC107979 Homo sapi
c	40	60	54.1	165649	9	AC103996	AC103996 Homo sapi
c	41	60	54.1	192826	9	AC090762	AC090762 Homo sapi
	42	59.8	53.9	138038	9	AL591435	AL591435 Human DNA
c	43	59.8	53.9	148352	2	AC025309	AC025309 Homo sapi
	44	59.8	53.9	171532	2	AC061993	AC061993 Homo sapi
c	45	59.8	53.9	173260	9	CNS07ECU	AL442183 Human chr

ALIGNMENTS

RESULT 1
AC012674/c
LOCUS AC012674 169620 bp DNA linear HTG 07-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP1-458H3, WORKING DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AC012674
VERSION AC012674.10 GI:9719580
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169620)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,

```
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
  REFERENCE 2  (bases 1 to 169620)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
  COMMENT   On Aug 7, 2000 this sequence version replaced gi:8705345.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMQG
                Center clone name: RP1-458H3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 139025 bases at least Q40
                Consensus quality: 154842 bases at least Q30
                Consensus quality: 159725 bases at least Q20
                Estimated insert size: 162720; sum-of-contigs estimation
                Estimated insert size: 171608; agarose-fp estimation
                Quality coverage: 3.9x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28689: contig of 28689 bp in length
            *    28690    28789: gap of unknown length
            *    28790    50832: contig of 22043 bp in length
            *    50833    50932: gap of unknown length
            *    50933    69144: contig of 18212 bp in length
            *    69145    69244: gap of unknown length
            *    69245    84204: contig of 14960 bp in length
            *    84205    84304: gap of unknown length
            *    84305    94667: contig of 10363 bp in length
            *    94668    94767: gap of unknown length
            *    94768   107261: contig of 12494 bp in length
            *   107262   107361: gap of unknown length
            *   107362   117550: contig of 10189 bp in length
            *   117551   117650: gap of unknown length
            *   117651   126939: contig of 9289 bp in length
            *   126940   127039: gap of unknown length
            *   127040   135040: contig of 8001 bp in length
            *   135041   135140: gap of unknown length
            *   135141   141639: contig of 6499 bp in length
            *   141640   141739: gap of unknown length
            *   141740   149558: contig of 7819 bp in length
            *   149559   149658: gap of unknown length
            *   149659   154562: contig of 4904 bp in length
            *   154563   154662: gap of unknown length
            *   154663   158987: contig of 4325 bp in length
            *   158988   159087: gap of unknown length
            *   159088   162376: contig of 3289 bp in length
            *   162377   162476: gap of unknown length
            *   162477   165191: contig of 2715 bp in length
            *   165192   165291: gap of unknown length
            *   165292   167173: contig of 1882 bp in length
            *   167174   167273: gap of unknown length
            *   167274   168393: contig of 1120 bp in length
            *   168394   168493: gap of unknown length
            *   168494   169620: contig of 1127 bp in length.
  FEATURES             Location/Qualifiers
       source          1..169620
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="3"
                       /clone="RP1-458H3"
  BASE COUNT    52024 a  33180 c  32128 g  50322 t   1966 others
  ORIGIN

    Query Match             100.0%;  Score 111;  DB 2;  Length 169620;
    Best Local Similarity   100.0%;  Pred. No. 6.5e-27;
    Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Db  87441 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 87382

  Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
            |||||||||||||||||||||||||||||||||||||||||||||||||||
  Db  87381 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 87331


  RESULT 2
  AC116565/c
  LOCUS       AC116565              199038 bp    DNA     linear   HTG 25-APR-2002
  DEFINITION  Homo sapiens chromosome 4 clone RP11-1263C18, WORKING DRAFT
              SEQUENCE, 6 unordered pieces.
  ACCESSION   AC116565
  VERSION     AC116565.3  GI:20304073
  KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
  SOURCE      Homo sapiens.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 199038)
    AUTHORS   Waterston,R.H.
    TITLE     The sequence of Homo sapiens clone
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 199038)
    AUTHORS   Waterston,R.H.
    TITLE     Direct Submission
    JOURNAL   Submitted (29-MAR-2002) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
  REFERENCE   3  (bases 1 to 199038)
    AUTHORS   Waterston,R.H.
    TITLE     Direct Submission
    JOURNAL   Submitted (25-APR-2002) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
  COMMENT     On Apr 25, 2002 this sequence version replaced gi:20128757.

              ------------------- Genome Center --------------------
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site:http://genome.wustl.edu/gsc/index.shtml
              Contact: submissions@watson.wustl.edu
```

```
            ---------- Project Information ----------
            Center project name: H_NH1263C18
            ----------------------------------------------------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 196905 bases at least Q40
            Consensus quality: 197086 bases at least Q30
            Consensus quality: 197234 bases at least Q20
            Insert size: 202000; agarose-fp
            Insert size: 198637; sum-of-contigs
            Quality coverage: 10.73 in Q20 bases; agarose-fp
            Quality coverage: 10.82 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3283: contig of 3283 bp in length
            *     3284     3383: gap of unknown length
            *     3384    12884: contig of 9501 bp in length
            *    12885    12984: gap of unknown length
            *    12985    29821: contig of 16837 bp in length
            *    29822    29921: gap of unknown length
            *    29922    74050: contig of 44129 bp in length
            *    74051    74150: gap of unknown length
            *    74151   197985: contig of 123835 bp in length
            *   197986   198085: gap of unknown length
            *   198086   199038: contig of 953 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199038
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-1263C18"
     misc_feature    1. .3283
                     /note="assembly_name:Contig18"
     misc_feature    3384. .12884
                     /note="assembly_name:Contig19"
     misc_feature    12985. .29821
                     /note="assembly_name:Contig20"
     misc_feature    29922. .74050
                     /note="assembly_name:Contig21"
     misc_feature    74151. .197985
                     /note="assembly_name:Contig22"
     misc_feature    198086. .199038
                     /note="assembly_name:Contig14"
BASE COUNT    53552 a  46319 c  46377 g  52290 t    500 others
ORIGIN

  Query Match             60.2%;  Score 66.8;  DB 2;  Length 199038;
  Best Local Similarity   81.9%;  Pred. No. 7.1e-12;
  Matches   77;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      18 CTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCT 77
           || | |||||   ||| |  | |||| ||  | ||||||||||||||||||||| |||
Db   58755 CTGGAAGGATGGATTTCCCACCTTGCCAGGGATCCTGGGGCCGGAGTATGTAAAACGCCT 58696

Qy      78 GGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||||||||| |||||||||||||| ||
Db   58695 GGGTCTCTGTGTGTGGCTGAGTGGCTGCTCTGCT 58662


RESULT 3
AC092586
LOCUS       AC092586              130636 bp    DNA     linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-30H15 from 2, complete sequence.
ACCESSION   AC092586 AC016389
VERSION     AC092586.2  GI:15638723
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130636)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130636)
  AUTHORS   Haglund,K., Kozlowicz,A., Elliott,G. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-30H15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 130636)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130636)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 130636)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:14916171.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0030H15
                Drafting Center: WIBR
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
```

```
          VECTOR:  pBACe3.6

          NEIGHBORING SEQUENCE INFORMATION:
          The clone sequenced to the left is RP11-111J6, 2000 bp overlap.
          Actual start of this clone is at base position 149730 of
          RP11-111J6; actual end is at base position 130636 of RP11-30H15.

          Polymorphisms have been identified between AC016389 and AC079148.

          The sequence of AC016389 has been incorporated into AC092586.
FEATURES             Location/Qualifiers
     source          1. .130636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-30H15"
                     /clone_lib="RPCI-11"
     repeat_region   245. .541
                     /rpt_family="Alu"
     repeat_region   1324. .1496
                     /rpt_family="MIR"
     repeat_region   1971. .2064
                     /rpt_family="L2"
     repeat_region   2910. .3139
                     /rpt_family="MIR"
     repeat_region   3175. .3200
                     /rpt_family="AT_rich"
     repeat_region   3326. .3387
                     /rpt_family="CR1"
     repeat_region   3562. .3697
                     /rpt_family="MIR"
     repeat_region   3936. .4102
                     /rpt_family="MIR"
     repeat_region   4319. .4484
                     /rpt_family="MIR"
     repeat_region   4958. .5268
                     /rpt_family="Alu"
     repeat_region   5242. .5276
                     /rpt_family="(TAAA)n"
     repeat_region   6051. .6112
                     /rpt_family="GA-rich"
     repeat_region   6902. .7270
                     /rpt_family="MaLR"
     repeat_region   7400. .7527
                     /rpt_family="L2"
     repeat_region   7537. .7680
                     /rpt_family="MIR"
     repeat_region   7782. .7826
                     /rpt_family="MIR"
     repeat_region   7851. .7910
                     /rpt_family="(TCCA)n"
     repeat_region   7898. .7953
                     /rpt_family="L2"
     repeat_region   7974. .8173
                     /rpt_family="MER1_type"
     repeat_region   8192. .8279
                     /rpt_family="L2"
     repeat_region   9263. .9292
                     /rpt_family="AT_rich"
     repeat_region   9315. .9354
                     /rpt_family="(TA)n"
     repeat_region   9373. .9529
                     /rpt_family="MIR"
     repeat_region   9584. .9612
                     /rpt_family="AT_rich"
     repeat_region   9601. .9954
                     /rpt_family="L1"
     repeat_region   10513. .11009
                     /rpt_family="L2"
     repeat_region   11300. .11740
                     /rpt_family="L2"
     repeat_region   11841. .12205
                     /rpt_family="MaLR"
     repeat_region   12208. .12897
                     /rpt_family="MaLR"
     repeat_region   12895. .14288
                     /rpt_family="L1"
     repeat_region   13495. .13523
                     /rpt_family="AT_rich"
     repeat_region   14321. .14437
                     /rpt_family="L1"
     repeat_region   14438. .14462
                     /rpt_family="(T)n"
     repeat_region   14747. .15888
                     /rpt_family="L1"
     repeat_region   15889. .15974
                     /rpt_family="(TA)n"
     repeat_region   15975. .15995
                     /rpt_family="(T)n"
     repeat_region   15989. .18191
                     /rpt_family="L1"
     repeat_region   16701. .16720
                     /rpt_family="(T)n"
     repeat_region   18193. .18704
                     /rpt_family="ERVL"
     repeat_region   18711. .18878
                     /rpt_family="L1"
     repeat_region   18879. .19175
                     /rpt_family="Alu"
     repeat_region   19176. .20870
                     /rpt_family="L1"
     repeat_region   21976. .22293
                     /rpt_family="L1"
     repeat_region   22376. .22535
                     /rpt_family="L1"
     repeat_region   22587. .22900
                     /rpt_family="L1"
     repeat_region   22906. .22981
                     /rpt_family="GA-rich"
     repeat_region   22969. .23321
                     /rpt_family="L1"
     repeat_region   23261. .23291
                     /rpt_family="AT_rich"
     repeat_region   23358. .23429
                     /rpt_family="L2"
     repeat_region   23718. .23827
                     /rpt_family="MIR"
     repeat_region   24408. .24560
                     /rpt_family="Alu"
     repeat_region   25160. .25330
                     /rpt_family="MIR"
     repeat_region   26042. .26343
                     /rpt_family="Alu"
     repeat_region   26324. .26343
                     /rpt_family="(A)n"

  Query Match             58.7%;  Score 65.2;  DB 9;  Length 130636;
  Best Local Similarity   74.5%;  Pred. No. 2.5e-11;
  Matches   82;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||  || ||||  ||||||||  || |      || |||||  |||| | ||||  | |
Db   20394 ATGTATGTATCTCATGCCTTGCCTGAGTTGCAGACACCTTTGTTGGGAATCCTGGAACTG 20453

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||| |||||||||||||||||||||||    |||||||| |
Db   20454 GAGTATGTAAAGCTCCTGGGTCTCTGTGTGTGCCTGTACAGCTGCTCTGC 20503


RESULT 4
AC022029/c
LOCUS       AC022029               140040 bp    DNA     linear   PRI 09-APR-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-93L14, complete sequence.
ACCESSION   AC022029
VERSION     AC022029.7  GI:19774403
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140040)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140040)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 140040)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 140040)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   5  (bases 1 to 140040)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Mar 28, 2002 this sequence version replaced gi:14718295.
            -------------- Genome Center
                Center: Genome Therapeutics Corporation
                Center code: GTC
                Web site: http://www.genomecorp.com/
                Contact: gtc-seqcenter@genomecorp.com
            -------------- Project Information
                Center project name: hg107

                IMPORTANT: This sequence is the entire
                insert of clone RP11-93L14.
                The true right end of clone RP11-316G7 is at
                2000 in this sequence.
            -------------- Summary Statistics
                Sequencing vector: N/A
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 990315
            --------------
FEATURES             Location/Qualifiers
     source          1. .140040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-93L14"
                     /clone_lib="RPCI-11"
BASE COUNT    48331 a  26787 c  24761 g  40161 t
ORIGIN

  Query Match             58.2%;  Score 64.6;  DB 9;  Length 140040;
  Best Local Similarity   73.9%;  Pred. No. 4e-11;
  Matches   82;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||  |||| ||   ||||||   || |  |  ||| ||||||  ||  | ||| ||| |
Db   94202 ATGTATGGACCTCCCGCCTTGACTGAGTTGTAGTCACCTTTGCCAGGGATCCTGTGGCTG 94143

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           | ||||||||| |||||||||||||||  | |||||||||||| || ||
Db   94142 GGGTATGTAAATCTCCTGGGTCTCTGTGCATACCTGAGTGGCTGCCCTGCT 94092


RESULT 5
AL355812
```

```
LOCUS       AL355812              147288 bp    DNA     linear   PRI 13-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-810O3 on chromosome Xq22.3-24,
            complete sequence.
ACCESSION   AL355812
VERSION     AL355812.23  GI:14456206
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147288)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 14, 2001 this sequence version replaced gi:14280423.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-810O3 is from the library RPCI-11.3 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-810O3 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-810O3 is at 1 in this sequence. The
            true left end of clone RP6-204F4 is at 147189 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .147288
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.3-24"
                     /clone="RP11-810O3"
                     /clone_lib="RPCI-11.3"
     repeat_region   1381. .1674
                     /note="AluJb repeat: matches 1. .296 of consensus"
     repeat_region   1757. .1937
                     /note="AluJo repeat: matches 122. .300 of consensus"
     repeat_region   2002. .2645
                     /note="L1M4 repeat: matches 2110. .2837 of consensus"
     repeat_region   2649. .3323
                     /note="L1MC4 repeat: matches 6404. .7210 of consensus"
     repeat_region   3811. .3878
                     /note="L1MEc repeat: matches 2099. .2180 of consensus"
     repeat_region   3882. .4044
                     /note="L1ME3A repeat: matches 5665. .5828 of consensus"
     repeat_region   4236. .4533
                     /note="L1ME3A repeat: matches 5257. .5563 of consensus"
     repeat_region   4724. .5174
                     /note="L1M4 repeat: matches 4562. .5026 of consensus"
     repeat_region   5401. .5663
```

```
                     /note="Charliela repeat: matches 1183. .1449 of consensus"
     repeat_region   5675. .5805
                     /note="L1PB3 repeat: matches 6014. .6145 of consensus"
     repeat_region   5807. .6234
                     /note="L1 repeat: matches 4970. .5403 of consensus"
     repeat_region   6251. .6998
                     /note="Charliela repeat: matches 404. .1196 of consensus"
     repeat_region   6999. .7036
                     /note="19 copies 2 mer tt 100% conserved"
     repeat_region   7244. .7864
                     /note="L1P repeat: matches 4160. .4781 of consensus"
     repeat_region   7865. .8274
                     /note="Charliela repeat: matches 6. .417 of consensus"
     repeat_region   8595. .8673
                     /note="MER58 repeat: matches 1. .83 of consensus"
     repeat_region   8689. .8885
                     /note="L1M4 repeat: matches 5270. .5464 of consensus"
     repeat_region   8942. .11399
                     /note="L1M4 repeat: matches 2661. .5149 of consensus"
     repeat_region   11442. .11832
                     /note="L1M4 repeat: matches 2153. .2568 of consensus"
     repeat_region   11837. .11904
                     /note="34 copies 2 mer tt 66% conserved"
     repeat_region   11917. .12078
                     /note="L1ME2 repeat: matches 5989. .6155 of consensus"
     repeat_region   12171. .13851
                     /note="L1MA9 repeat: matches 4533. .6275 of consensus"
     repeat_region   13877. .14584
                     /note="L1MA4 repeat: matches 5516. .6298 of consensus"
     repeat_region   14582. .16120
                     /note="L1M1 repeat: matches 1441. .2965 of consensus"
     repeat_region   16157. .16267
                     /note="L1M2 repeat: matches 1666. .1777 of consensus"
     repeat_region   16278. .17065
                     /note="L1ME1 repeat: matches 5224. .6053 of consensus"
     repeat_region   17066. .17368
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   17369. .17599
                     /note="L1ME1 repeat: matches 5010. .5224 of consensus"
     repeat_region   17654. .19512
                     /note="L1MA1 repeat: matches 4389. .6262 of consensus"
     repeat_region   19745. .20100
                     /note="L1MC2 repeat: matches 5470. .5826 of consensus"
     repeat_region   20094. .20104
                     /note="L1 repeat: matches 4404. .4413 of consensus"
     repeat_region   20105. .20666
                     /note="L1PA6 repeat: matches 5582. .6143 of consensus"
     repeat_region   20667. .21348
                     /note="L1 repeat: matches 3734. .4404 of consensus"
     repeat_region   21349. .21633
                     /note="AluSq repeat: matches 1. .291 of consensus"
     repeat_region   21634. .22811
                     /note="L1 repeat: matches 2511. .3734 of consensus"
     repeat_region   22853. .23268
                     /note="L1M4 repeat: matches 3293. .3703 of consensus"
     repeat_region   23269. .23420
                     /note="L1PA14 repeat: matches 6004. .6149 of consensus"
     repeat_region   23421. .24030
                     /note="L1M4 repeat: matches 2665. .3293 of consensus"
     repeat_region   24021. .24198
                     /note="L1M4 repeat: matches 2459. .2637 of consensus"
     repeat_region   24207. .24312
                     /note="L1M4 repeat: matches -3. .103 of consensus"
     repeat_region   24336. .24651
                     /note="L1MEc repeat: matches 2291. .2280 of consensus"
     repeat_region   24895. .25759
                     /note="L1MA9 repeat: matches 5440. .6304 of consensus"
     repeat_region   25760. .26060
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   26061. .26155
                     /note="L1MA9 repeat: matches 5346. .5440 of consensus"
     repeat_region   26213. .26559
                     /note="MLT1A1 repeat: matches 1. .354 of consensus"
     repeat_region   26561. .29502
                     /note="L1MA9 repeat: matches 2349. .5253 of consensus"
     repeat_region   29486. .29729
                     /note="L1MD3 repeat: matches 7485. .7734 of consensus"
     repeat_region   29737. .30026
                     /note="L1M4 repeat: matches 2689. .2989 of consensus"
     repeat_region   30048. .30693
                     /note="L1M4 repeat: matches 2187. .2831 of consensus"
     repeat_region   30716. .31150
                     /note="L1PB1 repeat: matches 5696. .6155 of consensus"
     repeat_region   31299. .31641
                     /note="L1MB7 repeat: matches 5814. .6166 of consensus"
     repeat_region   31664. .34500
                     /note="L1MB8 repeat: matches 3262. .6168 of consensus"
     repeat_region   34545. .34894
                     /note="L1MB8 repeat: matches 2920. .3279 of consensus"
     repeat_region   34901. .35618
                     /note="L1MA8 repeat: matches 5564. .6291 of consensus"
     repeat_region   35619. .36010
                     /note="L1MA8 repeat: matches 5204. .5597 of consensus"
     repeat_region   36008. .36242
                     /note="L1M4 repeat: matches 2694. .2931 of consensus"
     repeat_region   36243. .36433
                     /note="L1PA13 repeat: matches 5960. .6149 of consensus"
     repeat_region   36495. .36721
                     /note="Tigger3(Golem) repeat: matches 2759. .3038 of
                     consensus"
     repeat_region   37064. .37632
                     /note="L1M4 repeat: matches 2110. .2695 of consensus"
     repeat_region   37639. .37766
                     /note="64 copies 2 mer tt 59% conserved"
     repeat_region   38204. .38632
                     /note="MER65A repeat: matches 1. .445 of consensus"
     repeat_region   38696. .38991
                     /note="L1M4c repeat: matches 1768. .1603 of consensus"
     repeat_region   38992. .39287
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   39288. .39456
                     /note="L1M4c repeat: matches 1608. .1769 of consensus"
     repeat_region   39501. .40086
                     /note="L1M4b repeat: matches -273. .334 of consensus"
     repeat_region   40087. .40139
                     /note="L1MB7 repeat: matches 5776. .5828 of consensus"
     repeat_region   40136. .41315
                     /note="L1M4 repeat: matches -5. .1210 of consensus"
     repeat_region   41348. .42700
                     /note="L1M4c repeat: matches 940. .2376 of consensus"
     repeat_region   42986. .43636
                     /note="L1M4 repeat: matches 3439. .4077 of consensus"
     repeat_region   43654. .44154
                     /note="L1ME2 repeat: matches 4128. .4602 of consensus"
     repeat_region   44155. .44465
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   44466. .45911

  Query Match           58.2%;  Score 64.6;  DB 9;  Length 147288;
  Best Local Similarity  73.9%;  Pred. No. 4e-11;
  Matches   82;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||   ||| ||   |||||||  || |      || |||||  |||  | ||||||
Db   50611 ATGTACGGACCTCCCGCCTTGCCTGAGTTGCAGACACCTTTGTTGGGGATCCTGGGGTCA 50670

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||| |||||||||||||||||||||||  |||||| | ||
Db   50671 GAGTATGTAAAGCTCCTGGGTCTCTGTGTGTGCCTGAGCAGCTGCTTTGCT 50721


RESULT 6
AL591048/c
LOCUS       AL591048               75270 bp    DNA     linear   PRI 09-AUG-2001
DEFINITION  Human DNA sequence from clone RP11-612M16 on chromosome 6, complete
            sequence.
```

```
ACCESSION   AL591048
VERSION     AL591048.7  GI:15147695
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75270)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 10, 2001 this sequence version replaced gi:15020550.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-612M16 is from the library RPCI-11.3 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-612M16 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-206F19 is at 73271 in this sequence.
            The true right end of clone RP11-82M9 is at 2000 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .75270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-612M16"
                     /clone_lib="RPCI-11.3"
     misc_feature    33923. .34247
                     /note="Sequence from reads from a short insert library
                     derived  from a single pUC clone.  Restriction digest data
                     confirm the assembly."
     misc_feature    46760. .46849
                     /note="Sequence from reads from a short insert library
                     derived  from a single pUC clone.  Restriction digest data
                     confirm the assembly."
     misc_feature    49353. .49464
                     /note="Sequence from reads from a short insert library
                     derived  from a single pUC clone.  Restriction digest data
                     confirm the assembly."
BASE COUNT    25749 a  14850 c  13893 g  20778 t
ORIGIN

  Query Match             57.3%;  Score 63.6;  DB 9;  Length 75270;
  Best Local Similarity   73.6%;  Pred. No. 8.7e-11;
  Matches   81;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
             |||  |  || || |||||||||  ||    |   ||| ||||| ||   | ||||||| |

Db    1900 ATGTGCAGACCTCTTGCCTTGCCTGAGCTGTAGACATATTTGCTGGATATCCTGGGGCTG 1841

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||  ||  ||||||||  |||  ||||||||||   ||||  |||  |
Db    1840 GAGTATGTAAAGCTTCTGGGTCTGTGTCTGTGCCTGAGCAGCTGTTCTGC 1791


RESULT 7
AL137847/c
LOCUS       AL137847               143372 bp    DNA     linear   PRI 16-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-439K3 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL137847
VERSION     AL137847.12  GI:16973786
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143372)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 17, 2001 this sequence version replaced gi:16408610.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-439K3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-439K3 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-439K3 is at 1 in this sequence. The
            true left end of clone RP11-344L17 is at 141373 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .143372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-439K3"
                     /clone_lib="RPCI-11.2"
BASE COUNT    42809 a  30224 c  28152 g  42187 t
ORIGIN

  Query Match             57.3%;  Score 63.6;  DB 9;  Length 143372;
  Best Local Similarity   83.7%;  Pred. No. 8.7e-11;
  Matches   72;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      25 GATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTC 84
```

```
          |||   ||| |  | ||||  ||  | ||||||||| |||||||||||||||||||||||||
Db  82342 GATAGATTTCCTGCCTTGCCAGGGATCCTGGGGCCAGAGTATGTAAAACTCCTGGGTCTC 82283

Qy     85 TGTGTGTGCCTGAGTGGCTGCTCTAC 110
          |||||||||||||||||||||||||
Db  82282 TGTGTGTGCCTGAGTGGCTGCTCTGC 82257


RESULT 8
AL137179/c
LOCUS       AL137179              153053 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Human DNA sequence from clone RP11-82M9 on chromosome 6, complete
            sequence.
ACCESSION   AL137179
VERSION     AL137179.14  GI:10443372
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153053)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:9931245.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-82M9 is from the library RPCI-11.1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-82M9 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-82M9 is at 153053 in this
            sequence. The true right end of clone RP1-279I9 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .153053
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-82M9"
                     /clone_lib="RPCI-11.1"
     misc_feature    102234. .102253
                     /note="Tandem repeat. Forced join. Less than 1000bp
                     missing according to restriction digests."
BASE COUNT    47748 a  31096 c  30495 g  43714 t
ORIGIN

  Query Match            57.3%;  Score 63.6;  DB 9;  Length 153053;
  Best Local Similarity  73.6%;  Pred. No. 8.7e-11;
  Matches   81;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          ||| |  || || |||||||||  ||    |  ||| |||||  ||   | |||||||
Db 152945 ATGTGCAGACCTCTTGCCTTGCCTGAGCTGTAGACATATTTGCTGGATATCCTGGGGCTG 152886

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
          |||||||||| || |||||||| ||| ||||||||||  |||| ||| |
Db 152885 GAGTATGTAAAGCTTCTGGGTCTGTGTCTGTGCCTGAGCAGCTGTTCTGC 152836


RESULT 9
AC023408
LOCUS       AC023408              153422 bp    DNA     linear   HTG 07-JUN-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-612M16 map 6, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC023408
VERSION     AC023408.2  GI:8316895
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153422)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-612M16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153422)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 7, 2000 this sequence version replaced gi:6970573.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6363
                Center clone name: 612_M_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 135238 bases at least Q40
                Consensus quality: 141969 bases at least Q30
                Consensus quality: 146244 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 151322; sum-of-contigs
```

```
            Quality coverage: 4.0 in Q20 bases; agarose-fp
            Quality coverage: 3.9 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 22 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1152: contig of 1152 bp in length
          *     1153 1252: gap of      100 bp
          *     1253     2252: contig of 1000 bp in length
          *     2253 2352: gap of      100 bp
          *     2353     3503: contig of 1151 bp in length
          *     3504 3603: gap of      100 bp
          *     3604     4658: contig of 1055 bp in length
          *     4659 4758: gap of      100 bp
          *     4759     6059: contig of 1301 bp in length
          *     6060 6159: gap of      100 bp
          *     6160     6832: contig of 673 bp in length
          *     6833 6932: gap of      100 bp
          *     6933     9547: contig of 2615 bp in length
          *     9548 9647: gap of      100 bp
          *     9648    11267: contig of 1620 bp in length
          *    11268 11367: gap of      100 bp
          *    11368    14730: contig of 3363 bp in length
          *    14731 14830: gap of      100 bp
          *    14831    18501: contig of 3671 bp in length
          *    18502 18601: gap of      100 bp
          *    18602    23876: contig of 5275 bp in length
          *    23877 23976: gap of      100 bp
          *    23977    28125: contig of 4149 bp in length
          *    28126 28225: gap of      100 bp
          *    28226    33384: contig of 5159 bp in length
          *    33385 33484: gap of      100 bp
          *    33485    37950: contig of 4466 bp in length
          *    37951 38050: gap of      100 bp
          *    38051    43278: contig of 5228 bp in length
          *    43279 43378: gap of      100 bp
          *    43379    49641: contig of 6263 bp in length
          *    49642 49741: gap of      100 bp
          *    49742    57239: contig of 7498 bp in length
          *    57240 57339: gap of      100 bp
          *    57340    65126: contig of 7787 bp in length
          *    65127 65226: gap of      100 bp
          *    65227    74642: contig of 9416 bp in length
          *    74643 74742: gap of      100 bp
          *    74743    86960: contig of 12218 bp in length
          *    86961 87060: gap of      100 bp
          *    87061   115886: contig of 28826 bp in length
          *   115887 115986: gap of      100 bp
          *   115987   153422: contig of 37436 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-612M16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1152
                     /note="assembly_fragment"
     misc_feature    1253. .2252
                     /note="assembly_fragment"
     misc_feature    2353. .3503
                     /note="assembly_fragment"
     misc_feature    3604. .4658
                     /note="assembly_fragment"
     misc_feature    4759. .6059
                     /note="assembly_fragment"
     misc_feature    6160. .6832
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    6933. .9547
                     /note="assembly_fragment"
     misc_feature    9648. .11267
                     /note="assembly_fragment"
     misc_feature    11368. .14730
                     /note="assembly_fragment"
     misc_feature    14831. .18501
                     /note="assembly_fragment"
     misc_feature    18602. .23876
                     /note="assembly_fragment"
     misc_feature    23977. .28125
                     /note="assembly_fragment"
     misc_feature    28226. .33384
                     /note="assembly_fragment"
     misc_feature    33485. .37950
                     /note="assembly_fragment"
     misc_feature    38051. .43278
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    43379. .49641
                     /note="assembly_fragment"
     misc_feature    49742. .57239
                     /note="assembly_fragment"
     misc_feature    57340. .65126
                     /note="assembly_fragment"
     misc_feature    65227. .74642
                     /note="assembly_fragment"
     misc_feature    74743. .86960
                     /note="assembly_fragment"
     misc_feature    87061. .115886
                     /note="assembly_fragment"
     misc_feature    115987. .153422
                     /note="assembly_fragment"
BASE COUNT    47446 a  29588 c  28599 g  45687 t   2102 others
ORIGIN

  Query Match             57.3%;  Score 63.6;  DB 2;  Length 153422;
  Best Local Similarity   73.6%;  Pred. No. 8.7e-11;
  Matches   81;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
             ||| |  || || ||||||||  ||    |   ||| ||||| ||   | |||||||
Db   49270 ATGTGCAGACCTCTTGCCTTGCCTGAGCTGTAGACATATTTGCTGGATATCCTGGGGCTG 49329

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||| || |||||||| ||| ||||||||||  |||| ||| |
Db   49330 GAGTATGTAAAGCTTCTGGGTCTGTGTCTGTGCCTGAGCAGCTGTTCTGC 49379


RESULT 10
AC109592/c
LOCUS       AC109592               86314 bp    DNA     linear   PRI 13-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-750O22 from 4, complete sequence.
ACCESSION   AC109592
VERSION     AC109592.3  GI:19310357
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86314)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 86314)
  AUTHORS   Buatsi,D. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-750O22
```

```
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 86314)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 86314)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 86314)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 86314)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 9, 2002 this sequence version replaced gi:18642965.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0750O22
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-510D4, 2000 bp overlap; the
            clone sequenced to the right is RP11-366I1, 2000 bp overlap.
            Actual start of this clone is at base position 149897 of
            RP11-510D4; actual end is at base position 44359 of RP11-366I1.
FEATURES             Location/Qualifiers
     source          1. .86314
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-750O22"
                     /clone_lib="RPCI-11"
     repeat_region   2. .160
                     /rpt_family="Alu"
     repeat_region   256. .655
                     /rpt_family="L1"
     repeat_region   954. .1182
                     /rpt_family="L1"
     repeat_region   1216. .1472
                     /rpt_family="L1"
     repeat_region   1563. .1876
                     /rpt_family="L1"
     repeat_region   2896. .3255
                     /rpt_family="MaLR"
     repeat_region   3256. .4669
                     /rpt_family="MaLR"
     repeat_region   4670. .4969
                     /rpt_family="MaLR"
     repeat_region   10088. .10184
                     /rpt_family="MaLR"
     repeat_region   11452. .11617
                     /rpt_family="CR1"
     repeat_region   12716. .12930
                     /rpt_family="CR1"
     repeat_region   13258. .13395
                     /rpt_family="MIR"
     repeat_region   14378. .14415
                     /rpt_family="AT_rich"
     repeat_region   14817. .14883
                     /rpt_family="Alu"
     misc_feature    15540. .15842
                     /note="match to EST AI023549 (NID:g3238593) ov79f12.s1"
     repeat_region   21639. .21945
                     /rpt_family="L1"
     repeat_region   23095. .23142
                     /rpt_family="(TG)n"
     repeat_region   24837. .24987
                     /rpt_family="ERVL"
     repeat_region   24994. .26519
                     /rpt_family="ERVL"
     repeat_region   26650. .26833
                     /rpt_family="ERVL"
     repeat_region   26834. .26918
                     /rpt_family="(TA)n"
     repeat_region   26938. .26973
                     /rpt_family="(GA)n"
     repeat_region   26979. .27252
                     /rpt_family="Alu"
     repeat_region   27546. .27618
                     /rpt_family="MaLR"
     repeat_region   27984. .28617
                     /rpt_family="L1"
     repeat_region   28624. .28942
                     /rpt_family="Alu"
     repeat_region   29073. .29150
                     /rpt_family="MER53"
     repeat_region   29081. .29158
                     /rpt_family="MER53"
     repeat_region   29199. .29562
                     /rpt_family="L1"
     repeat_region   29563. .30039
                     /rpt_family="L1"
     repeat_region   30313. .31555
                     /rpt_family="L1"
     repeat_region   31554. .31738
                     /rpt_family="L1"
     repeat_region   31739. .31786
                     /rpt_family="AT_rich"
     repeat_region   34024. .34352
                     /rpt_family="Alu"
     repeat_region   38180. .38776
                     /rpt_family="L1"
```

```
     repeat_region   38867. .39139
                     /rpt_family="L1"
     repeat_region   39172. .39448
                     /rpt_family="L1"
     repeat_region   39523. .39821
                     /rpt_family="Alu"
     repeat_region   40157. .40231
                     /rpt_family="AT_rich"
     repeat_region   40341. .40597
                     /rpt_family="L1"
     repeat_region   40604. .40670
                     /rpt_family="(TG)n"
     repeat_region   40758. .40966
                     /rpt_family="L1"
     repeat_region   41763. .41802
                     /rpt_family="AT_rich"
     repeat_region   42632. .42724
                     /rpt_family="(CATATA)n"
     repeat_region   42858. .42895
                     /rpt_family="L1"
     repeat_region   42901. .43787
                     /rpt_family="L1"
     repeat_region   43786. .45120
                     /rpt_family="L1"
     repeat_region   45123. .45419
                     /rpt_family="MaLR"
     repeat_region   45531. .47107
                     /rpt_family="MaLR"
     repeat_region   47108. .47262
                     /rpt_family="MaLR"
     repeat_region   47263. .48876
                     /rpt_family="L1"
     repeat_region   51158. .51280
                     /rpt_family="AcHobo"
     repeat_region   51281. .51586
                     /rpt_family="L1"

  Query Match             57.1%;  Score 63.4;  DB 9;  Length 86314;
  Best Local Similarity   82.0%;  Pred. No. 1e-10;
  Matches   73;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      23 AGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTC 82
           |||||   | |||  | |||  |||| | ||||||| ||||||||||||||| ||||||
Db   32282 AGGATGGATCTTCTGCCTTGGCGGGAATCCTGGGGCTGGAGTATGTAAAACTCTTGGGTC 32223

Qy      83 TCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||||||  |||||||||||| ||
Db   32222 TCTGTGTGTGCCCCAGTGGCTGCTCTGCT 32194

RESULT 11
AC021378/c
LOCUS       AC021378              159475 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-26G6, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC021378
VERSION     AC021378.4  GI:8072570
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159475)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-26G6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159475)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C.; Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7637277.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4633
                Center clone name: 26_G_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150152 bases at least Q40
                Consensus quality: 155379 bases at least Q30
                Consensus quality: 157415 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 158575; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2315: contig of 2315 bp in length
            *     2316 2415: gap of      100 bp
            *     2416     9041: contig of 6626 bp in length
            *     9042 9141: gap of      100 bp
            *     9142    17359: contig of 8218 bp in length
            *    17360 17459: gap of      100 bp
            *    17460    26067: contig of 8608 bp in length
            *    26068 26167: gap of      100 bp
            *    26168    40876: contig of 14709 bp in length
            *    40877 40976: gap of      100 bp
            *    40977    56845: contig of 15869 bp in length
            *    56846 56945: gap of      100 bp
            *    56946    74524: contig of 17579 bp in length
            *    74525 74624: gap of      100 bp
            *    74625    94662: contig of 20038 bp in length
            *    94663 94762: gap of      100 bp
            *    94763   126925: contig of 32163 bp in length
            *   126926 127025: gap of      100 bp
            *   127026   159475: contig of 32450 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-26G6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2315
                     /note="assembly_fragment
```

```
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2416. .9041
                     /note="assembly_fragment"
     misc_feature    9142. .17359
                     /note="assembly_fragment"
     misc_feature    17460. .26067
                     /note="assembly_fragment"
     misc_feature    26168. .40876
                     /note="assembly_fragment"
     misc_feature    40977. .56845
                     /note="assembly_fragment"
     misc_feature    56946. .74524
                     /note="assembly_fragment"
     misc_feature    74625. .94662
                     /note="assembly_fragment"
     misc_feature    94763. .126925
                     /note="assembly_fragment"
     misc_feature    127026. .159475
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
BASE COUNT    52037 a  27831 c  28355 g  50348 t    904 others
ORIGIN

  Query Match             57.1%;  Score 63.4;  DB 2;  Length 159475;
  Best Local Similarity   82.0%;  Pred. No. 1e-10;
  Matches   73;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      23 AGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTC 82
           |||||   | |||  | |||  |||| | ||||||| |||||||||||||||| ||||||
Db   45765 AGGATGGATCTTCTGCCTTGGCGGGAATCCTGGGGCTGGAGTATGTAAAACTCTTGGGTC 45706

Qy      83 TCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||||  |||||||||||| ||
Db   45705 TCTGTGTGTGCCCCAGTGGCTGCTCTGCT 45677


RESULT 12
AL157937/c
LOCUS       AL157937              153783 bp    DNA     linear   PRI 04-APR-2001
DEFINITION  Human DNA sequence from clone RP11-408N14 on chromosome 9 Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL157937
VERSION     AL157937.21  GI:11121006
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153783)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:11071307.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-408N14 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-408N14 is at 1 in this sequence.
            The true left end of clone RP11-441I5 is at 153684 in this
            sequence. The true right end of clone RP11-145E5 is at 29400 in
            this sequence. This sequence was finished as follows unless
            otherwise noted: all regions were either double-stranded or
            sequenced with an alternate chemistry or covered by high quality
            data (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats; all
            regions were covered by at least one plasmid subclone or more than
            one M13 subclone; and the assembly was confirmed by restriction
            digest. RP11-408N14 is from the library RPCI-11.2 constructed by
            the group of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .153783
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-408N14"
                     /clone_lib="RPCI-11.2"
     repeat_region   26. .191
                     /note="MIR repeat: matches 21. .190 of consensus"
     repeat_region   499. .941
                     /note="L1M4 repeat: matches 4921. .5372 of consensus"
     repeat_region   1151. .1457
                     /note="AluJb repeat: matches 1. .304 of consensus"
     repeat_region   1671. .2296
                     /note="L1MC5 repeat: matches 7249. .7897 of consensus"
     repeat_region   2401. .2642
                     /note="L1MC5 repeat: matches 6983. .7227 of consensus"
     repeat_region   3314. .3359
                     /note="23 copies 2 mer tg 93% conserved"
     repeat_region   3769. .3991
                     /note="MIR repeat: matches 5. .241 of consensus"
     misc_feature    5093. .5394
                     /note="match: GSS: Em:AQ823856"
     repeat_region   5306. .5391
                     /note="L2 repeat: matches 2164. .2262 of consensus"
     repeat_region   5557. .5881
                     /note="MLT2FA repeat: matches 5. .324 of consensus"
     misc_feature    5607. .5973
                     /note="match: GSS: Em:AQ137344"
     misc_feature    6531. .7177
                     /note="match: GSS: Em:B91342"
     misc_feature    6538. .6935
                     /note="match: GSS: Em:AQ217550"
     repeat_region   6611. .6910
                     /note="AluJo repeat: matches 1. .297 of consensus"
     repeat_region   7761. .7811
                     /note="L2 repeat: matches 2651. .2694 of consensus"
     repeat_region   7812. .8120
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   8121. .8174
                     /note="L2 repeat: matches 2576. .2651 of consensus"
     misc_feature    complement(8940. .9357)
                     /note="match: GSS: Em:AQ443387"
     repeat_region   8950. .9125
                     /note="MIR repeat: matches 20. .204 of consensus"
     repeat_region   9105. .9285
                     /note="MIR repeat: matches 29. .213 of consensus"
     misc_feature    9360. .9878
                     /note="match: GSS: Em:AQ771822"
     repeat_region   9753. .9824
                     /note="36 copies 2 mer tg 79% conserved"
     repeat_region   9869. .9980
                     /note="MIR repeat: matches 139. .262 of consensus"
     repeat_region   11255. .17292
                     /note="L1PA5 repeat: matches 76. .6143 of consensus"
     repeat_region   17910. .18096
                     /note="L1MC4 repeat: matches 7641. .7835 of consensus"
```

```
repeat_region   18520. .20148
                /note="L1PA12 repeat: matches -1421. .1009 of consensus"
repeat_region   20127. .21370
                /note="L1PA11 repeat: matches 921. .2161 of consensus"
repeat_region   21372. .25281
                /note="L1PA11 repeat: matches 2262. .6165 of consensus"
repeat_region   25293. .25713
                /note="L1PA10 repeat: matches 5655. .6083 of consensus"
repeat_region   26120. .26412
                /note="AluSx repeat: matches 3. .295 of consensus"
misc_feature    26198. .26577
                /note="match: GSS: Em:AQ147782"
repeat_region   26660. .26962
                /note="AluY repeat: matches 1. .300 of consensus"
misc_feature    complement(27055. .27448)
                /note="match: STS: Em:HSJ37G11"
repeat_region   27407. .31040
                /note="L1PA12 repeat: matches 2505. .6164 of consensus"
repeat_region   31031. .31138
                /note="L1P repeat: matches 1541. .1650 of consensus"
repeat_region   31144. .31497
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   31498. .33130
                /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                consensus"
repeat_region   33131. .33479
                /note="THE1B repeat: matches 1. .359 of consensus"
repeat_region   33485. .35130
                /note="L1PA13 repeat: matches -248. .1557 of consensus"
repeat_region   36624. .36669
                /note="23 copies 2 mer ta 76% conserved"
repeat_region   36680. .36897
                /note="L1MC4 repeat: matches 6404. .6623 of consensus"
repeat_region   37435. .37508
                /note="37 copies 2 mer ag 75% conserved"
repeat_region   38923. .39151
                /note="Charlie4 repeat: matches 231. .1926 of consensus"
misc_feature    complement(39281. .39766)
                /note="match: GSS: Em:AQ894025"
repeat_region   40317. .40540
                /note="MIR repeat: matches 20. .234 of consensus"
repeat_region   40603. .40726
                /note="L2 repeat: matches 2587. .2710 of consensus"
repeat_region   43871. .43952
                /note="L2 repeat: matches 2673. .2750 of consensus"
repeat_region   43962. .44259
                /note="AluY repeat: matches 1. .299 of consensus"
repeat_region   45076. .45280
                /note="MIR repeat: matches 2. .205 of consensus"
repeat_region   45413. .45588
                /note="MIR repeat: matches 1. .205 of consensus"
misc_feature    complement(46149. .46586)
                /note="match: GSS: Em:AQ343245"
repeat_region   47817. .47938
                /note="L2 repeat: matches 2627. .2749 of consensus"
repeat_region   48349. .48579
                /note="MLT1I repeat: matches 1. .220 of consensus"
repeat_region   50262. .50664
                /note="L2 repeat: matches 2339. .2749 of consensus"
repeat_region   50674. .52080
                /note="L1M1 repeat: matches -1389. .30 of consensus"
repeat_region   52084. .52853
                /note="L1MA2 repeat: matches 5516. .6308 of consensus"
repeat_region   54498. .60638
                /note="L1PA4 repeat: matches 1. .6146 of consensus"
repeat_region   61367. .61560
                /note="MLT1F repeat: matches 346. .541 of consensus"
repeat_region   62305. .62430
                /note="63 copies 2 mer tt 59% conserved"
repeat_region   63089. .63629
                /note="L1PA12 repeat: matches 10. .540 of consensus"
repeat_region   63630. .63887
                /note="AluY repeat: matches 7. .263 of consensus"
repeat_region   63888. .64059
                /note="AluSg repeat: matches 134. .303 of consensus"
repeat_region   64060. .64352
                /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region   64353. .64485
                /note="AluSg repeat: matches 1. .134 of consensus"
repeat_region   64486. .64834
                /note="L1PA12 repeat: matches 539. .-291 of consensus"
repeat_region   64816. .64938
                /note="L1PA10 repeat: matches 1761. .1897 of consensus"
repeat_region   64939. .65350
                /note="LTR14B repeat: matches 1. .608 of consensus"
repeat_region   65355. .67266
                /note="L1PA10 repeat: matches 1894. .3818 of consensus"
repeat_region   67317. .69162
                /note="L1PA10 repeat: matches 3793. .5643 of consensus"
repeat_region   69122. .69167
                /note="L1P4 repeat: matches 5616. .5662 of consensus"
repeat_region   69216. .69754
                /note="L1PA10 repeat: matches 5610. .6165 of consensus"
repeat_region   70283. .70389

  Query Match             56.8%;  Score 63;  DB 9;  Length 153783;
  Best Local Similarity   73.0%;  Pred. No. 1.4e-10;
  Matches   81;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          |||    ||  |  |||||||   || |      || |||||| |||  | |||||||||
Db  18987 ATGTACAGACATCCTGCCTTGTCTGAGTTGCAGACACCTTTGCTGGGGATCCTGGGGCCG 18928

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
          |||||||||| ||||||||||||||||  | ||||||  |||||||| ||
Db  18927 GAGTATGTAAAGCTCCTGGGTCTCTGTGCATACCTGAGCAGCTGCTCTGCT 18877


RESULT 13
AC021025/c
ID   AC021025   standard; DNA; HTG; 123779 BP.
XX
AC   AC021025;
XX
SV   AC021025.9
XX
DT   14-JAN-2000 (Rel. 62, Created)
DT   08-AUG-2000 (Rel. 64, Last updated, Version 11)
XX
DE   Homo sapiens chromosome 3 clone RP11-79K17, WORKING DRAFT SEQUENCE, 10
DE   unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-123779
RA   Muzny D.M., Adams C., Bailey M., Barbaria J., Blankenburg K., Bodota B.,
RA   Bouck J., Bowie S., Brooks A., Buhay C., Bunac C., Burkett C., Burrows J.,
RA   Carter M., Chacko J., Chen Z., Cox C., David R., Delgado O., Deshazo D.,
RA   Ding Y., Domah-Rashid N., Dugan-Rocha S., Durbin K.J., Fernandez C.,
RA   Ferraguto D., Forcum-Tansey J., Frantz P., Ganesh R., Gorrell J.H.,
RA   Gorrell L.L., Guevara W., Harris K., Hernandez J., Hodgson A., Hogues M.,
RA   Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M.,
RA   Kelly S., Kondejewski N., Kong Y., Kovar C., Leal B., Li Z., Lichtarge O.,
RA   Liu J., Liu W., Logan O., Lozado R.J., Lu J., Lucier R., Martin R.,
RA   Martinez C., McLeod M.P., Mei G., Morgan M., Morris S., Nash S., Nelson A.,
RA   Nguyen R., Nguyen N., Nguyen S., Oswal G., Parish B., Paxton S., Payton B.,
RA   Perez L., Pu L.L., Quiles M., Reiter D., Rives M., Samuel S., Say J.,
RA   Scherer S., Shah E., Shen H., Simon M., Sparks A., Stamps A., Sucgang R.,
RA   Tabor P., Taylor T., Vasquez L., Vinson R., Vo Q., Wahbah M.,
RA   Watlington S., Weinstock G., Weinstock I.R., Williamson A., Worley K.,
RA   Wren J., Wrensford G., Yu W., Zhou X., Nelson D., Gibbs R.;
```

```
RT   "Direct Submission";
RL   Unpublished.
XX
RN   [2]
RP   1-123779
RA   Worley K.C.;
RT   ;
RL   Submitted (12-JAN-2000) to the EMBL/GenBank/DDBJ databases.
RL   Human Genome Sequencing Center, Department of Molecular and Human Genetics,
RL   Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
XX
CC   On Aug 7, 2000 this sequence version replaced gi:8700003.
CC   -------------- Genome Center
CC   Center: Baylor College of Medicine
CC   Center code: BCM
CC   Web site: http://www.hgsc.bcm.tmc.edu/
CC   Contact: hgsc-help@bcm.tmc.edu
CC   -------------- Project Information
CC   Center project name: HMXB
CC   Center clone name: RP11-79K17
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; L08821
CC   Chemistry: Dye-terminator Big Dye: 100% of reads
CC   Assembly program: Phrap; version 0.990329
CC   Consensus quality: 112332 bases at least Q40
CC   Consensus quality: 117272 bases at least Q30
CC   Consensus quality: 119789 bases at least Q20
CC   Estimated insert size: 121963; sum-of-contigs estimation
CC   Quality coverage: 0x in Q20 bases; agarose-fp estimation
CC   Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation
CC   --------------
CC   * NOTE: Estimated insert size may differ from sequence length
CC   *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 10 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1    35931: contig of 35931 bp in length
CC   *    35932    36031: gap of unknown length
CC   *    36032    60754: contig of 24723 bp in length
CC   *    60755    60854: gap of unknown length
CC   *    60855    71667: contig of 10813 bp in length
CC   *    71668    71767: gap of unknown length
     *    71768    84826: contig of 13059 bp in length
     *    84827    84926: gap of unknown length
     *    84927    96721: contig of 11795 bp in length
CC   *    96722    96821: gap of unknown length
CC   *    96822   105594: contig of 8773 bp in length
CC   *   105595   105694: gap of unknown length
CC   *   105695   112907: contig of 7213 bp in length
CC   *   112908   113007: gap of unknown length
CC   *   113008   120198: contig of 7191 bp in length
CC   *   120199   120298: gap of unknown length
CC   *   120299   122220: contig of 1922 bp in length
CC   *   122221   122320: gap of unknown length
CC   *   122321   123779: contig of 1459 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .123779
FT                   /chromosome="3"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-79K17"
XX
SQ   Sequence 123779 BP; 37802 A; 24017 C; 23437 G; 37612 T; 911 other;

  Query Match            56.2%;  Score 62.4;  DB 30;  Length 123779;
  Best Local Similarity  81.8%;  Pred. No. 2.2e-10;

  Matches   72;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           ||||   | | |  | |||| | || | ||||||  |||||||| ||||||||||||||||
Db    3128 GGATGGATCTCCCACCTTGCTGAGAATCCTGGGGGTGGAGTATGCAAAACTCCTGGGTCT 3069

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||||||||||||||||||| ||
Db    3068 CTGTGTGTGCCTGAGTGGCTGCTCTGCT 3041


RESULT 14
AC076969
LOCUS       AC076969              128118 bp    DNA     linear   HTG 15-OCT-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K12, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC076969
VERSION     AC076969.6  GI:16117967
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128118)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128118)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:10047573.
            -------------- Genome Center
```

```
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
        ---------- Project Information
            Center project name: HBRJ
            Center clone name: RP11-79K12
        ---------- Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 111307 bases at least Q40
            Consensus quality: 117905 bases at least Q30
            Consensus quality: 121494 bases at least Q20
            Estimated insert size: 122854; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
        --------------
        * NOTE: Estimated insert size may differ from sequence length
        *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 14 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1    20354: contig of 20354 bp in length
        *    20355    20454: gap of unknown length
        *    20455    33071: contig of 12617 bp in length
        *    33072    33171: gap of unknown length
        *    33172    46935: contig of 13764 bp in length
        *    46936    47035: gap of unknown length
        *    47036    57066: contig of 10031 bp in length
        *    57067    57166: gap of unknown length
        *    57167    70378: contig of 13212 bp in length
        *    70379    70478: gap of unknown length
        *    70479    81786: contig of 11308 bp in length
        *    81787    81886: gap of unknown length
        *    81887    91428: contig of 9542 bp in length
        *    91429    91528: gap of unknown length
        *    91529   100455: contig of 8927 bp in length
        *   100456   100555: gap of unknown length
        *   100556   107879: contig of 7324 bp in length
        *   107880   107979: gap of unknown length
        *   107980   114149: contig of 6170 bp in length
        *   114150   114249: gap of unknown length
        *   114250   120186: contig of 5937 bp in length
        *   120187   120286: gap of unknown length
        *   120287   123236: contig of 2950 bp in length
        *   123237   123336: gap of unknown length
        *   123337   125606: contig of 2270 bp in length
        *   125607   125706: gap of unknown length
        *   125707   128118: contig of 2412 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128118
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K12"
BASE COUNT    35433 a  25053 c  23792 g  42367 t   1473 others
ORIGIN

  Query Match            56.2%;  Score 62.4;  DB 2;  Length 128118;
  Best Local Similarity  81.8%;  Pred. No. 2.2e-10;
  Matches   72;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           ||||   | | |  | |||| | || | ||||||  |||||||| ||||||||||||||
Db   55061 GGATGGATCTCCCACCTTGCTGAGAATCCTGGGGGTGGAGTATGCAAAACTCCTGGGTCT 55120

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||||||||||||||||||||| ||
Db   55121 CTGTGTGTGCCTGAGTGGCTGCTCTGCT 55148


RESULT 15
AC121249/c
LOCUS       AC121249              128583 bp    DNA     linear   PRI 01-JUN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-79K17, complete sequence.
ACCESSION   AC121249 AC021025
VERSION     AC121249.1  GI:20806313
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 128583)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 16, 2002 this sequence version replaced gi:9719675.
            ---------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            ---------- Project Information
                Center project name: chr-3
                Center clone name: RP11-79K17 (bc0196)
            ---------- Summary Statistics
                Sequencing vector: plasmid; 45% of reads
                Sequencing vector: M13; L08821; 55% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 128548 bases at least Q40
                Consensus quality: 128580 bases at least Q30
                Consensus quality: 128581 bases at least Q20
                Insert size: 128583; sum-of-contigs
                Quality coverage: 13.4x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  RP11-147N17 (UWGC:bc0267) AC104300, 22116-bp overlap
                3':  RP11-391P4 (UWGC:bc0402) AC104303, 114502-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
```

covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
12049	12136	4077	4013	8696	8902
579	<800	2067	2145	6	<800
1056	1065	9698	9932	1323	1337
13085	13126	3729	3734	1024	961
10944	10836	1632	1616	8421	8408
581	<800	8069	8139	1114	1126
882	887	1507	1485	3538	3534
8435	8467	11939	11871	6051	6022
3924	4103	133	<800	5129	5048
9842	9759	2135	2145	58	<800
4401	4377	78	<800	1895	1903
4458	4377	5873	5861	799	<800
4587	4522	896	902	858	885
6048	5947	3312	3356	931	885
1395	1385	1505	1485	859	885
3059	3071	6694	6716	3934	3936
3911	3861	10032	9932	1065	1058
138	<800	2168	2145	2629	2723
50	<800	1794	1785	859	885
582	<800	227	<800	6855	6888
2225	2229	4002	4013	7821	7819
984	967	2714	2726	7210	7201
887	887	5628	5624	6992	6888
1042	1065	743	758	1204	1224
2783	2808	3045	2963	2722	2723
7074	7158	1288	1262	321	<800
4087	3861	967	973	1066	1058
7976	7926	461	<800	470	<800
1233	1199	4649	4618	3055	3072
3875	3861	5289	5417	10961	10848
27	<800	1098	1095	6725	6888
1137	1065	798	809	1896	1903
750	749	111	<800	8682	8666
214	<800	7644	7609	4438	4342
836	887	10084	9932	514	<800
1670	1651	3641	3589	3077	3072
2650	2609	7555	7609	1371	1337
1890	1872			4185	4087
4615	4522			719	<800
1321	1273			4130	4087
				367	<800
				2979	3072
				333	<800

```
FEATURES             Location/Qualifiers
     source          1. .128583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-79K17"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    37777 a  24752 c  24789 g  41265 t
ORIGIN

  Query Match             56.2%;  Score 62.4;  DB 9;  Length 128583;
  Best Local Similarity   81.8%;  Pred. No. 2.2e-10;
  Matches   72;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           ||||   | | |  | |||| | || | ||||||  |||||||| ||||||||||||||
Db   27328 GGATGGATCTCCCACCTTGCTGAGAATCCTGGGGGTGGAGTATGCAAAACTCCTGGGTCT 27269

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||||||||||||||||||||| ||
Db   27268 CTGTGTGTGCCTGAGTGGCTGCTCTGCT 27241


Search completed: April 25, 2003, 02:17:54
Job time : 1357 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 00:52:59 ; Search time 150 Seconds
(without alignments)
1666.479 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

1 number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	100.0	447	21	AAC03794	Human secreted pro
	2	111	100.0	447	21	AAZ42680	Human 5' EST isola
c	3	58.2	52.4	349	22	AAF65888	Novel human polynu
c	4	58.2	52.4	570	22	ABA63453	Human foetal liver
c	5	58.2	52.4	570	22	ABA30652	Probe #9118 for ge
c	6	58.2	52.4	570	22	AAK11985	Human brain expres
c	7	58.2	52.4	570	22	AAK37688	Human bone marrow
c	8	58.2	52.4	570	22	AAI18447	Probe #8380 for ge
c	9	58.2	52.4	570	22	AAI43563	Probe #12249 used
c	10	58.2	52.4	570	24	ABS11680	Human genome-deriv
c	11	56.4	50.8	128600	24	ABK83461	Human cDNA differe
c	12	55	49.5	1982	21	AAC68089	Human secreted pro
	13	52.8	47.6	609	22	AAH98980	Human EST-derived
c	14	50.8	45.8	1299	23	AAS91827	DNA encoding novel
c	15	46	41.4	807	23	AAS64647	DNA encoding novel
c	16	46	41.4	843	23	AAS79254	DNA encoding novel
c	17	44	39.6	1996	22	ABA18896	Human nervous syst
c	18	43.8	39.5	1909	23	AAS88434	DNA encoding novel
c	19	40.2	36.2	240	24	ABN25478	Human ORFX polynuc
c	20	39	35.1	660	23	AAS73441	DNA encoding novel
	21	38.8	35.0	162	22	ABA75762	Human foetal liver
	22	38.8	35.0	519	22	ABA63320	Human foetal liver
c	23	36.2	32.6	2787	23	AAS68019	DNA encoding novel
c	24	36.2	32.6	2787	23	AAS69828	DNA encoding novel
c	25	35.2	31.7	1173	23	AAS69200	DNA encoding novel
c	26	35.2	31.7	1173	23	AAS82303	DNA encoding novel
c	27	35.2	31.7	1173	23	AAS86064	DNA encoding novel
c	28	35.2	31.7	1291	23	AAS77610	DNA encoding novel
c	29	35.2	31.7	2097	23	AAS77443	DNA encoding novel
	30	35.2	31.7	2569	23	AAS77685	DNA encoding novel
	31	35.2	31.7	11534	24	ABL32342	Human immune syste
c	32	34.2	30.8	1278	23	AAS72474	DNA encoding novel
c	33	33.4	30.1	396	22	AAK74465	Human immune/haema
	34	32.4	29.2	447	22	AAL16101	Human breast cance
	35	32.4	29.2	447	22	AAL24945	Human breast cance
	36	31.8	28.6	349980	22	AAF86431	Pyrococcus abyssi
	37	31.8	28.6	349980	22	AAH41223	Pyrococcus abyssi
	38	31	27.9	826	22	AAH04786	Human cDNA clone (
	39	30.8	27.7	3659	7	AAN60204	Interferon-pseudo-
c	40	30.6	27.6	352	22	AAK59083	Human immune/haema
c	41	29.2	26.3	369	24	ABK64361	Human benign prost
c	42	29.2	26.3	369	24	ABL61736	Colon adenocarcino
c	43	29.2	26.3	369	24	ABL65612	Lung cancer relate
c	44	29.2	26.3	369	24	ABL66228	Lung cancer relate
	45	29.2	26.3	1314	22	AAS31530	Human DNA for a no

ALIGNMENTS

RESULT 1
AAC03794
ID AAC03794 standard; cDNA; 447 BP.
XX
AC AAC03794;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3792.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG03788.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3792; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 111; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.4e-27;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 2
AAZ42680
ID AAZ42680 standard; cDNA; 447 BP.
XX
AC AAZ42680;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:439.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
OS Homo sapiens.
XX
PN WO9953051-A2.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
DR P-PSDB; AAY65066.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 402; 837pp; English.
XX
CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 447 BP; 74 A; 105 C; 125 G; 138 T; 5 other;

Query Match 100.0%; Score 111; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.4e-27;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 51 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 110

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
|||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 161

RESULT 3
AAF65888/c
ID AAF65888 standard; cDNA; 349 BP.
XX
AC AAF65888;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 1644.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX

PS Claim 9; Page 779; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 349 BP; 90 A; 93 C; 102 G; 64 T; 0 other;

Query Match 52.4%; Score 58.2; DB 22; Length 349;
Best Local Similarity 70.3%; Pred. No. 7.1e-10;
Matches 78; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
Db 134 ATGTGCAGACCTCTTGCCTTGCCTGAGTTGCAAACATGTTTGTTGGAAATCCCAGTGCCG 75

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
Db 74 GAGTACGTAAAGCACCTGGGTCTCTGTGCATGCCTGAACAGCTGCTCCCCT 24

RESULT 4
ABA63453/c
ID ABA63453 standard; DNA; 570 BP.
XX
AC ABA63453;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #11758.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 11758; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Query Match 52.4%; Score 58.2; DB 22; Length 570;
Best Local Similarity 79.3%; Pred. No. 8.2e-10;
Matches 69; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
Db 566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy 84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
Db 506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480

RESULT 5
ABA30652/c
ID ABA30652 standard; DNA; 570 BP.
XX
AC ABA30652;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #9118 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 9118; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

```
CC   congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             52.4%;  Score 58.2;  DB 22;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 8.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480


RESULT 6
AAK11985/c
ID   AAK11985 standard; DNA; 570 BP.
XX
AC   AAK11985;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 11976.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 11976; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             52.4%;  Score 58.2;  DB 22;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 8.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480


RESULT 7
AAK37688/c
ID   AAK37688 standard; DNA; 570 BP.
XX
AC   AAK37688;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 12245.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 12245; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             52.4%;  Score 58.2;  DB 22;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 8.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480
```

```
RESULT 8
AAI18447/c
ID   AAI18447 standard; DNA; 570 BP.
XX
AC   AAI18447;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #8380 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
     30-JAN-2001; 2001WO-US00670.

     04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 8380; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
     Note: The sequence data for this patent did not form part of the printed
     specification, but was obtained in electronic format directly from WIPO
     at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             52.4%;  Score 58.2;  DB 22;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 8.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480


RESULT 9
AAI43563/c
ID   AAI43563 standard; DNA; 570 BP.
XX
AC   AAI43563;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #12249 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 12249; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

  Query Match             52.4%;  Score 58.2;  DB 22;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 8.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480


RESULT 10
ABS11680/c
ID   ABS11680 standard; DNA; 570 BP.
XX
AC   ABS11680;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human genome-derived single exon probe from lung SEQ ID No 11671.
XX
KW   Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
```

XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 1; SEQ ID No 11671; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 570 BP; 169 A; 154 C; 156 G; 91 T; 0 other;

Query Match 52.4%; Score 58.2; DB 24; Length 570;
Best Local Similarity 79.3%; Pred. No. 8.2e-10;

Matches 69; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| |||||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480
```

RESULT 11
ABK83461/c
ID ABK83461 standard; cDNA; 128600 BP.
XX
AC ABK83461;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #32.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 32; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for

CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 128600 BP; 36139 A; 24970 C; 26316 G; 41175 T; 0 other;

[redacted]ery Match 50.8%; Score 56.4; DB 24; Length 128600;
[redacted]st Local Similarity 70.8%; Pred. No. 1.6e-08;
Matches 75; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
Db 126322 ATGTGTGGACCTCCCACCTTGACTTCATTGCAGACGCATTTGTTGGGGCTCCTGGGGCTG 126263

Qy 61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCT 106
Db 126262 GAGTATGTAAAGCTCCTGGGTCTCTGTGCATGCCTGAGCAGCTGCT 126217

RESULT 12
AAC68089/c
ID AAC68089 standard; cDNA; 1982 BP.
XX
AC AAC68089;
XX
DT 20-FEB-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #9.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
[redacted] Homo sapiens.
XX
PN WO200058335-A1.
XX
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US07534.
XX
PR 26-MAR-1999; 99US-0126598.
PR 22-DEC-1999; 99US-0171504.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611702/58.
DR P-PSDB; AAB37356.
XX
PT Nucleic acids encoding human secreted proteins, used to treat, prevent,
PT ameliorate or diagnose conditions such as cancer, and autoimmune
PT diseases e.g. arthritis -
XX
PS Claim 1; Pages 321-322; 387pp; English.
XX

CC The invention relates to the isolation of genes AAC68081-C68127 encoding
CC 47 human secreted proteins AAB37348-B37394. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC68072) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 1982 BP; 549 A; 513 C; 471 G; 448 T; 1 other;

Query Match 49.5%; Score 55; DB 21; Length 1982;
Best Local Similarity 70.9%; Pred. No. 1.4e-08;
Matches 73; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 9 ATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGT 68
Db 1163 ATCTCCCGCTTTGCAGGAGTTGCAGTTACTTTTGCTGGGAAACCCAGGGCCGGAGTATGT 1104

Qy 69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
Db 1103 AAAGCTGCTGGGTCTCTGCACGTGCTCAAGCGGCTACTCTGCT 1061

RESULT 13
AAH98980
ID AAH98980 standard; cDNA; 609 BP.
XX
AC AAH98980;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 837.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR P-PSDB; AAM24321.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -

XX
PS Claim 1; Page 694; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 609 BP; 99 A; 162 C; 173 G; 175 T; 0 other;

Query Match 47.6%; Score 52.8; DB 22; Length 609;
Best Local Similarity 71.9%; Pred. No. 5.5e-08;
Matches 69; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 15 TGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACT 74
272 TGCCTTGCCTGAGATGCCATCACCTTTCTTGGGGATCCTAAGGCTGGAGTATGTAAAGCT 331

Qy 75 CCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
Db 332 CCTGGGTCTCTGTATGTGCCTGAGCACCGGTTCTTC 367

RESULT 14
AAS91827/c
ID AAS91827 standard; cDNA; 1299 BP.
XX
AC AAS91827;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27631.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27640.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 27631; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1299 BP; 399 A; 349 C; 271 G; 280 T; 0 other;

Query Match 45.8%; Score 50.8; DB 23; Length 1299;
Best Local Similarity 71.3%; Pred. No. 3.2e-07;
Matches 67; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 17 CCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCC 76
Db 897 CCTTGCCTGAGTTGCAGCCACATTTGTTGGGGATCCTGGGGCTGGAGTATGTAAAGCTCC 838

Qy 77 TGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
Db 837 CAGGTCTTTGTGCATGCCGGAGCAGCTGCTCTGC 804

RESULT 15
AAS64647/c
ID AAS64647 standard; cDNA; 807 BP.
XX
AC AAS64647;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #451.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG00460.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 451; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 807 BP; 224 A; 213 C; 190 G; 180 T; 0 other;

Query Match 41.4%; Score 46; DB 23; Length 807;
Best Local Similarity 65.7%; Pred. No. 1.1e-05;
Matches 67; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy       9 ATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGT 68
           | ||| ||| ||||||||| |       |  ||||  ||||   | || || |||||||
Db     158 ACCTTCTGCTTTGCAGGAGTTGCAGATACTTTTGTTGGGAAGACAGGAGCTGGAGTATCT 99

Qy      69 AAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           ||| ||||   || ||||| |||| |||||||  | |  | |
Db      98 AAAGCTCCCAAGTATCTGTATGTGTCTGAGTGTTTTCCTTGC 57
```

Search completed: April 25, 2003, 01:55:16
Job time : 166 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 01:47:50 ; Search time 1042 Seconds
(without alignments)
1725.239 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

1 number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	100.0	470	17	AQ770688	AQ770688 HS_5368_B
	2	99.4	89.5	525	17	AQ165256	AQ165256 HS_3025_B
c	3	63.4	57.1	519	17	AQ186743	AQ186743 HS_3109_B
	4	63	56.8	416	17	AQ182486	AQ182486 HS_3077_A
c	5	63	56.8	867	12	BG675164	BG675164 602621440
	6	61.4	55.3	516	17	AQ457001	AQ457001 HS_5151_A
c	7	58.8	53.0	529	17	AQ881246	AQ881246 HS_5137_B
c	8	58.4	52.6	597	17	AQ506984	AQ506984 RPCI-11-3
	9	58.2	52.4	416	17	B75884	B75884 RPCI11-12L1
c	10	58.2	52.4	530	17	AQ193128	AQ193128 HS_3060_B
	11	57.8	52.1	448	17	AQ702903	AQ702903 HS_5452_B
	12	57.8	52.1	769	17	AQ899390	AQ899390 HS_5234_A
	13	57.6	51.9	413	17	AQ266658	AQ266658 RPCI11-73
	14	57.2	51.5	410	17	AQ442274	AQ442274 HS_5137_A
	15	57.2	51.5	453	17	AQ437684	AQ437684 HS_5137_A
c	16	57.2	51.5	635	17	AQ390599	AQ390599 CITBI-E1-
c	17	56.8	51.2	519	17	AQ139984	AQ139984 HS_3106_A
	18	56.6	51.0	491	17	AQ186162	AQ186162 HS_3077_A
	19	56.6	51.0	628	17	AQ237815	AQ237815 RPCI11-70
c	20	56.6	51.0	864	17	AQ739814	AQ739814 HS_5505_A
	21	56	50.5	546	17	AQ637256	AQ637256 RPCI-11-4
	22	55.6	50.1	360	17	AQ207172	AQ207172 HS_3239_B
	23	55.6	50.1	513	17	AQ455447	AQ455447 HS_5153_A
	24	55.6	50.1	541	17	AQ765592	AQ765592 HS_5348_B
	25	55.6	50.1	546	17	AQ541696	AQ541696 RPCI-11-3
	26	55.6	50.1	666	17	AQ427698	AQ427698 CITBI-E1-
	27	55.6	50.1	732	17	AQ506035	AQ506035 RPCI-11-3
c	28	55.4	49.9	580	17	AQ532835	AQ532835 RPCI-11-3
	29	55.2	49.7	695	17	AG179297	AG179297 Pan trogl
	30	55	49.5	423	17	AQ564722	AQ564722 HS_5361_A
c	31	55	49.5	482	17	AQ320567	AQ320567 RPCI11-99
	32	55	49.5	549	17	AQ540868	AQ540868 RPCI-11-3
	33	55	49.5	551	17	AQ569689	AQ569689 HS_5333_B
c	34	55	49.5	563	17	AQ420187	AQ420187 RPCI-11-1
	35	55	49.5	580	17	AQ316169	AQ316169 RPCI11-10
	36	55	49.5	634	17	AG160901	AG160901 Pan trogl
	37	55	49.5	659	17	AG151043	AG151043 Pan trogl
c	38	55	49.5	723	17	AQ386439	AQ386439 RPCI11-15
	39	54.8	49.4	292	17	AQ508480	AQ508480 RPCI-11-2
	40	54.6	49.2	483	17	AQ668395	AQ668395 HS_5414_A
c	41	54.6	49.2	653	17	AG143347	AG143347 Pan trogl
	42	54.6	49.2	683	17	AG091225	AG091225 Pan trogl
	43	54.4	49.0	503	17	AQ155611	AQ155611 HS_3124_A
c	44	54.4	49.0	708	17	AG092466	AG092466 Pan trogl
	45	54.2	48.8	672	17	AG051939	AG051939 Pan trogl

ALIGNMENTS

RESULT 1
AQ770688
LOCUS AQ770688 470 bp DNA linear GSS 28-JUL-1999
DEFINITION HS_5368_B2_C08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=944 Col=16 Row=F, DNA sequence.
ACCESSION AQ770688
VERSION AQ770688.1 GI:5648804
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

```
          (pieter@dejong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
          or from Resear h Genetics (info@resgen.com). BAC end Web Server:
          http://www.htsc.washington.edu
          Plate: 944  row: F  column: 16
          Seq primer: SP6
          Class: BAC ends
          High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=944 Col=16 Row=F"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       83 a    112 c    131 g    141 t      3 others
ORIGIN

  Query Match             100.0%;  Score 111;  DB 17;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-23;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 162

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 213


RESULT 2
AQ165256
LOCUS       AQ165256               525 bp    DNA     linear   GSS 16-OCT-1998
DEFINITION  HS_3025_B2_G06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3025 Col=12 Row=N, DNA sequence.
ACCESSION   AQ165256
VERSION     AQ165256.1  GI:3563451
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    RENCE   1  (bases 1 to 525)
    THORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3025  row: N  column: 12
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3025 Col=12 Row=N"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      102 a    139 c    137 g    143 t      4 others
ORIGIN

  Query Match             89.5%;  Score 99.4;  DB 17;  Length 525;
  Best Local Similarity   92.8%;  Pred. No. 7.6e-20;
  Matches  103;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
             ||||| |||||||| ||||||||||    | | ||||||||||||||||||||||||||||
Db      66 ATGGGCGGATCTTTCGCCTTGCAGGAAATTATNTCATCTTTGCAGGGACTTCTGGGGCCG 125

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
             |||||||||||||||||| ||||||||||||||||||||||||||||||
Db     126 GAGTATGTAAAACTCCTGNGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 176


RESULT 3
AQ186743/c
LOCUS       AQ186743               519 bp    DNA     linear   GSS 01-NOV-1998
DEFINITION  HS_3109_B1_A06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3109 Col=11 Row=B, DNA sequence.
ACCESSION   AQ186743
VERSION     AQ186743.1  GI:3586185
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3109  row: B  column: 11
            Class: BAC ends
            High quality sequence stop: 519.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3109 Col=11 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      144 a    155 c    122 g     93 t      5 others
ORIGIN

  Query Match             57.1%;  Score 63.4;  DB 17;  Length 519;
  Best Local Similarity   82.0%;  Pred. No. 6.3e-09;
  Matches   73;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      23 AGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTC 82
           |||||   | |||  | |||  |||| | |||||||  ||||||||||||||| ||||||
Db     398 AGGATGGATCTTCTGCCTTGGCGGGAATCCTGGGGCTGGAGTATGTAAAACTCTTGGGTC 339

Qy      83 TCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||||||  |||||||||||| ||
Db     338 TCTGTGTGTGCCCCAGTGGCTGCTCTGCT 310
```

```
RESULT 4
AQ182486
LOCUS       AQ182486                416 bp    DNA     linear   GSS 28-OCT-1998
DEFINITION  HS_3077_A1_G04_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3077 Col=7 Row=M, DNA sequence.
ACCESSION   AQ182486
VERSION     AQ182486.1  GI:3579853
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3077  row: M  column: 7
            Class: BAC ends
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3077 Col=7 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       68 a    111 c    121 g    113 t      3 others
ORIGIN

  Query Match             56.8%;  Score 63;  DB 17;  Length 416;
  Best Local Similarity   73.0%;  Pred. No. 8e-09;
  Matches   81;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||    ||  |  |||||||   || |      || |||||| |||  | |||||||||
Db      96 ATGTACAGACATCCTGCCTTGTCTGAGTTGCAGACACCTTTGCTGGGGATCCTGGGGCCG 155

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||||| ||||||||||||||||  | ||||||  |||||||| ||
Db     156 GAGTATGTAAAGCTCCTGGGTCTCTGTGCATACCTGAGCAGCTGCTCTGCT 206


RESULT 5
BG675164/c
LOCUS       BG675164                867 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602621440F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4755296 5',
            mRNA sequence.
ACCESSION   BG675164
VERSION     BG675164.1  GI:13906560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10617  row: o  column: 09
            High quality sequence start: 5
            High quality sequence stop: 857.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4755296"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      285 a    208 c    217 g    157 t
ORIGIN

  Query Match             56.8%;  Score 63;  DB 12;  Length 867;
  Best Local Similarity   73.0%;  Pred. No. 9.2e-09;
  Matches   81;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||  |||| ||   ||||||   || |  |  ||| ||||||  ||  | ||| ||| |
Db     636 ATGTATGGACCTCCCGCCTTGACTGAGTTGTAGTCACCTTTGCCAGGGATCCTGTGGCTG 577

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           | |||||||| ||| ||||||||||||  | ||||||||||||| || ||
Db     576 GGGTATGTAAATCTCTTGGGTCTCTGTGCATACCTGAGTGGCTGCCCTGCT 526


RESULT 6
AQ457001
LOCUS       AQ457001                516 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  HS_5151_A1_G09_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=727 Col=17 Row=M, DNA sequence.
ACCESSION   AQ457001
VERSION     AQ457001.1  GI:4635878
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
```

```
            Plate: 727  row: M  column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 516.
FEATURES              Location/Qualifiers
     source           1. .516
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=727 Col=17 Row=M"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT       86 a    126 c    145 g    151 t      8 others
ORIGIN

  Query Match             55.3%;  Score 61.4;  DB 17;  Length 516;
  Best Local Similarity   72.1%;  Pred. No. 2.6e-08;
  Matches   80;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
           |||    ||  |  |||||||   || |      || ||| || |||  | |||||||||
Db     130 ATGTACAGACATCCTGCCTTGTCTGAGTTGCAAACACCTTCGCTGGGGATCCTGGGGCCG 189

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||| ||||||||||||||||  | ||||||  |||||||| ||
Db     190 GAGTATGTAAAGCTCCTGGGTCTCTGTGCATACCTGAGCAGCTGCTCTGCT 240


RESULT 7
AQ881246/c
LOCUS       AQ881246                529 bp    DNA     linear   GSS 09-NOV-1999
DEFINITION  HS_5137_B1_F08_T7 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=8905 Col=15 Row=L, DNA sequence.
ACCESSION   AQ881246
VERSION     AQ881246.1  GI:6312713
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 8905  row: L  column: 15
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 529.
FEATURES              Location/Qualifiers
     source           1. .529
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=8905 Col=15 Row=L"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      137 a    147 c    121 g    117 t      7 others
ORIGIN

  Query Match             53.0%;  Score 58.8;  DB 17;  Length 529;
  Best Local Similarity   80.2%;  Pred. No. 1.6e-07;
  Matches   69;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      25 GATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTC 84
           |||    | | |  | |  | |||| ||| ||||| ||||||||||||||||||||||||
Db     417 GATGAATCTCCCACCTCACCGGGATTTCCGGGGCTGGAGTATGTAAAACTCCTGGGTCTC 358

Qy      85 TGTGTGTGCCTGAGTGGCTGCTCTAC 110
           ||||||||| |||| |||||||||
Db     357 TGTGTGTGCATGAGCGGCTGCTCTGC 332


RESULT 8
AQ506984/c
LOCUS       AQ506984                597 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  RPCI-11-316J23.TV RPCI-11 Homo sapiens genomic clone RPCI-11-316J23
            , DNA sequence.
ACCESSION   AQ506984
VERSION     AQ506984.1  GI:4711731
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-316J23.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .597
                      /organism="Homo sapiens"
                      /db_xref="GDB:7621198"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-316J23"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      158 a    161 c    159 g    119 t
ORIGIN

  Query Match             52.6%;  Score 58.4;  DB 17;  Length 597;
```

```
  Best Local Similarity   77.2%;  Pred. No. 2.1e-07;
  Matches   71;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      19 TTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 78
           | |  |||| ||| | |  | |||  ||   | |||| ||| |||| ||||||||||||
Db     348 TGGATGGATACTTCTCCCACCTTGATGGAGATCCTGGAGCCAGAGTGTGTAAAACTCCTA 289

Qy      79 GGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||||| |||||||| |
Db     288 GGTCTCTGTGTGTGCCTGAGCGGCTGCTCTGC 257


RESULT 9
B75884
LOCUS       B75884                   416 bp    DNA     linear   GSS 08-APR-1999
DEFINITION  RPCI11-12L15.TP RPCI-11 Homo sapiens genomic clone RPCI-11-12L15,
            DNA sequence.
ACCESSION   B75884
VERSION     B75884.1  GI:2771571
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-12L15.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Homo sapiens"
                     /db_xref="GDB:7504502"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-12L15"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       76 a    103 c    120 g    117 t
ORIGIN

  Query Match             52.4%;  Score 58.2;  DB 17;  Length 416;
  Best Local Similarity   70.3%;  Pred. No. 2.3e-07;
  Matches   78;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
             |||  |||| |||   ||||||  || |     |||||||||   ||  | |||||  |
Db     113 ATGTATGGACCTTCGACCTTGCCTGAGTTGCAGTCATCTTTGTCAGGGATCCTGGGCTCA 172

Qy      61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||  ||||||  |||||  ||||||  || |||||||||  |||||  || ||
Db     173 GAGTCTGTAAAGTTCCTGAGTCTCTCTGCGTGCCTGAGCAGCTGCACTGCT 223
```

```
RESULT 10
AQ193128/c
LOCUS       AQ193128                 530 bp    DNA     linear   GSS 04-NOV-1998
DEFINITION  HS_3060_B1_E08_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3060 Col=15 Row=J, DNA sequence.
ACCESSION   AQ193128
VERSION     AQ193128.1  GI:3591750
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3060  row: J  column: 15
            Class: BAC ends
            High quality sequence stop: 530.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3060 Col=15 Row=J"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      154 a    152 c    120 g     99 t      5 others
ORIGIN

  Query Match             52.4%;  Score 58.2;  DB 17;  Length 530;
  Best Local Similarity   83.5%;  Pred. No. 2.4e-07;
  Matches   66;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      33 TTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTCTGTGTGTG 92
           | |  | |||| |||| ||| || |||||||  ||||||||||||||||||||||||||
Db     419 TCCCGCCTTGCTGGGATTTCAGGAGCCGGAGACTGTAAAACTCCTGGGTCTCTGTGTGTG 360

Qy      93 CCTGAGTGGCTGCTCTACT 111
           | |||||| ||||||| ||
Db     359 CTTGAGTGACTGCTCTGCT 341


RESULT 11
AQ702903
LOCUS       AQ702903                 448 bp    DNA     linear   GSS 07-JUL-1999
DEFINITION  HS_5452_B1_H02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1028 Col=3 Row=P, DNA sequence.
ACCESSION   AQ702903
VERSION     AQ702903.1  GI:5412329
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
```

```
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1028  row: P  column: 3
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 448.
    URES              Location/Qualifiers
     source           1. .448
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=1028 Col=3 Row=P"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT       72 a    104 c    126 g    140 t      6 others
ORIGIN

  Query Match             52.1%;  Score 57.8;  DB 17;  Length 448;
  Best Local Similarity   78.2%;  Pred. No. 3.1e-07;
  Matches   68;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           ||  ||   | ||| || |||| | |  |||| |||| | | || |||||||| ||||||
Db     110 GGTTTGATCTTCTTCCTTGCGGNGGATTCTTGGGCTGAANTAGGTAAAACTTCTGGGTTT 169

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           || ||||||||||||||||||||||| |
Db     170 CTCTGTGTGCCTGAGTGGCTGCTCTGC 196


RESULT 12
AQ899390
LOCUS       AQ899390               769 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_5234_A1_E09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9002 Col=17 Row=I, DNA sequence.
ACCESSION   AQ899390
VERSION     AQ899390.1  GI:6355580
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9002  row: I  column: 17
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 769.
FEATURES              Location/Qualifiers
     source           1. .769
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=9002 Col=17 Row=I"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      145 a    163 c    196 g    265 t
ORIGIN

  Query Match             52.1%;  Score 57.8;  DB 17;  Length 769;
  Best Local Similarity   76.3%;  Pred. No. 3.4e-07;
  Matches   71;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      19 TTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 78
           | |  ||||   ||||   ||||||  |   |||||||||  || |||||||| |||||
Db     115 TGGATGGATGGATTTTTCACTTTGCTAGTGATTCTGGGGCAAGAATATGTAAAATTCCTG 174

Qy      79 GGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||  ||||| ||||||||||||||| |||||
Db     175 GGTCCTTGTGTTTGCCTGAGTGGCTGCACTACT 207


RESULT 13
AQ266658
LOCUS       AQ266658               413 bp    DNA     linear   GSS 27-APR-1999
DEFINITION  RPCI11-73J15.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-73J15,
            DNA sequence.
ACCESSION   AQ266658
VERSION     AQ266658.1  GI:3794262
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-73J15.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /db_xref="GDB:7527878"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-73J15"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       75 a    108 c    118 g    112 t
ORIGIN

  Query Match             51.9%;  Score 57.6;  DB 17;  Length 413;
  Best Local Similarity   78.4%;  Pred. No. 3.5e-07;
  Matches   69;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           ||||   | | |  | ||||| |||   ||| ||| |||||||| ||||||||||||||
Db     104 GGATGGATCTCCCACCTTGCAAGGAAACCTGCGGCTGGAGTATGCAAAACTCCTGGGTCT 163

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||||||| | || ||||||||||| ||
Db     164 CTGTGTGTGGCCGACTGGCTGCTCTGCT 191


RESULT 14
AQ442274
LOCUS       AQ442274                410 bp    DNA     linear   GSS 31-MAR-1999
DEFINITION  HS_5137_A1_F12_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=713 Col=23 Row=K, DNA sequence.
ACCESSION   AQ442274
VERSION     AQ442274.1  GI:4553613
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 713  row: K  column: 23
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=713 Col=23 Row=K"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       80 a    111 c    111 g    107 t      1 others
ORIGIN

  Query Match             51.5%;  Score 57.2;  DB 17;  Length 410;
  Best Local Similarity   79.1%;  Pred. No. 4.6e-07;
  Matches   68;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      25 GATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTC 84
           |||  || |||  | |||| |    | |  ||||| ||||||| |||| ||||||||||
Db     111 GATGGTTCTTCTGCCTTGCTGCTGATCCCAGGGCCAGAGTATGCAAAATTCCTGGGTCTC 170

Qy      85 TGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||||||||| |
Db     171 TGTGTGTGCCTGAGTGGCTGCTCTGC 196


RESULT 15
AQ437684
LOCUS       AQ437684                453 bp    DNA     linear   GSS 31-MAR-1999
DEFINITION  HS_5137_A2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=713 Col=12 Row=O, DNA sequence.
ACCESSION   AQ437684
VERSION     AQ437684.1  GI:4549023
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 713  row: O  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=713 Col=12 Row=O"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       84 a    127 c    117 g    124 t      1 others
ORIGIN

  Query Match             51.5%;  Score 57.2;  DB 17;  Length 453;
  Best Local Similarity   79.1%;  Pred. No. 4.7e-07;
```

Matches 68; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy      25 GATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTC 84
           |||  || |||  | |||| |    | |  ||||| ||||||| |||| ||||||||||| 
Db     114 GATGGTTCTTCTGCCTTGCTGCTGATCCCAGGGCCAGAGTATGCAAAATTCCTGGGTCTC 173

Qy      85 TGTGTGTGCCTGAGTGGCTGCTCTAC 110
           ||||||||||||||||||||||||| |
Db     174 TGTGTGTGCCTGAGTGGCTGCTCTGC 199
```

Search completed: April 25, 2003, 02:28:38
Job time : 1045 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 01:55:25 ; Search time 39 Seconds
(without alignments)
872.849 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

1 number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	27.6	24.9	80246	4	US-09-078-294-4	Sequence 4, Appli
	2	27.6	24.9	80595	4	US-09-078-294-3	Sequence 3, Appli
	3	26.2	23.6	175	1	US-08-222-177A-4	Sequence 4, Appli
	4	26.2	23.6	1146	2	US-08-666-367B-4	Sequence 4, Appli
c	5	26.2	23.6	1146	4	US-09-143-438-4	Sequence 4, Appli
c	6	26.2	23.6	13104	4	US-08-961-527-34	Sequence 34, Appl
c	7	26	23.4	430	4	US-09-397-787-254	Sequence 254, App
c	8	25.4	22.9	11492	4	US-08-991-840A-1	Sequence 1, Appli
c	9	24.8	22.3	1371	2	US-08-428-713-1	Sequence 1, Appli
c	10	24.8	22.3	1371	3	US-08-904-179-1	Sequence 1, Appli
c	11	24.8	22.3	1374	2	US-08-428-713-9	Sequence 9, Appli
c	12	24.8	22.3	1374	3	US-08-904-179-9	Sequence 9, Appli
c	13	24.6	22.2	882	4	US-09-556-877-136	Sequence 136, App
c	14	24.6	22.2	882	4	US-09-620-412C-136	Sequence 136, App
	15	24.6	22.2	2407	4	US-09-370-807-7	Sequence 7, Appli
	16	24.6	22.2	2407	4	US-09-921-259-7	Sequence 7, Appli
c	17	24.6	22.2	2511	4	US-09-422-869-19	Sequence 19, Appl
	18	24.4	22.0	361	4	US-09-385-982-26	Sequence 26, Appl
	19	24.4	22.0	1506	2	US-08-206-790A-22	Sequence 22, Appl
	20	24.4	22.0	1506	5	PCT-US95-02943-22	Sequence 22, Appl
c	21	24.2	21.8	2347	1	US-08-145-681-3	Sequence 3, Appli
c	22	24.2	21.8	2347	1	US-08-453-703-3	Sequence 3, Appli
c	23	24.2	21.8	2347	2	US-08-456-106-3	Sequence 3, Appli
c	24	24.2	21.8	2347	3	US-08-456-108-3	Sequence 3, Appli
c	25	24.2	21.8	2347	4	US-09-265-577-3	Sequence 3, Appli
c	26	24	21.6	528	4	US-09-134-001C-2094	Sequence 2094, Ap
	27	24	21.6	4088	2	US-08-317-310A-1	Sequence 1, Appli
	28	24	21.6	4088	5	PCT-US95-13041-1	Sequence 1, Appli
c	29	24	21.6	19446	4	US-08-961-527-51	Sequence 51, Appl
c	30	23.8	21.4	3831	4	US-08-961-527-291	Sequence 291, App
c	31	23.8	21.4	7542	4	US-09-734-030-3	Sequence 3, Appli
	32	23.8	21.4	13158	2	US-08-687-080-105	Sequence 105, App
	33	23.8	21.4	35100	1	US-08-306-691B-19	Sequence 19, Appl
	34	23.8	21.4	35100	5	PCT-US93-06251-19	Sequence 19, Appl
	35	23.8	21.4	87350	3	US-08-781-891-79	Sequence 79, Appl
	36	23.8	21.4	87543	4	US-09-791-211-3	Sequence 3, Appli
c	37	23.6	21.3	15567	4	US-09-627-376-3	Sequence 3, Appli
	38	23.4	21.1	560	3	US-09-059-369-18	Sequence 18, Appl
	39	23.4	21.1	598	4	US-08-998-416-1076	Sequence 1076, Ap
	40	23.4	21.1	694	4	US-08-998-416-710	Sequence 710, App
	41	23.4	21.1	1477	4	US-09-123-030-9	Sequence 9, Appli
	42	23.4	21.1	2090	3	US-09-059-369-1	Sequence 1, Appli
c	43	23.4	21.1	3093	1	US-08-252-966B-19	Sequence 19, Appl
c	44	23.4	21.1	5253	2	US-08-290-731C-3	Sequence 3, Appli
	45	23.2	20.9	401	4	US-09-221-298-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 24.9%; Score 27.6; DB 4; Length 80246;
Best Local Similarity 58.5%; Pred. No. 5.3;
Matches 48; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy      23 AGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTC 82
           ||  |||| | ||| |||  | ||| | |  |  | |    |||||||     | |  |
Db    6333 AGTTTTCTGTGTCAACTTGACTGGGCCATGGGATGTCCAGATATGTAATTAAACAGTATT 6392

Qy      83 TCTGTGTGTGCCTGAGTGGCTG 104
           |||| ||||  ||| | || ||
Db    6393 TCTGGGTGTTTCTGTGAGGGTG 6414
```

RESULT 2
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA

```
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

  Query Match             24.9%;  Score 27.6;  DB 4;  Length 80595;
  Best Local Similarity   58.5%;  Pred. No. 5.3;
  Matches   48;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      23 AGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTC 82
           ||  |||| | ||| |||  | ||| | |  |  | |    |||||||     | |  |
Db    6618 AGTTTTCTGTGTCAACTTGACTGGGCCATGGGATGTCCAGATATGTAATTAAACAGTATT 6677

Qy      83 TCTGTGTGTGCCTGAGTGGCTG 104
           |||| ||||  ||| | || ||
Db    6678 TCTGGGTGTTTCTGTGAGGGTG 6699


RESULT 3
US-08-222-177A-4
; Sequence 4, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
;   APPLICANT:  Weber, James L.
;   TITLE OF INVENTION:  LENGTH POLYMORPHISMS IN
;   TITLE OF INVENTION:  (dC-dA)n.(dG-dT)n SEQUENCES AND METHODS OF USING SAME
;   NUMBER OF SEQUENCES:  460
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  DeWitt Ross & Stevens, S.C.
;     STREET:  8000 Excelsior Drive, Suite 401
;     CITY:  Madison
;     STATE:  Wisconsin
;     COUNTRY:  USA
;     ZIP:  53717-1914
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/222,177A
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/341,562
;     FILING DATE:  21-APR-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Sara, Charles S.
;     REGISTRATION NUMBER:  30,492
      REFERENCE/DOCKET NUMBER:  09865.601
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (608) 831-2100
;     TELEFAX:  (608) 831-2106
;     TELEX:
; INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  175 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   ORIGINAL SOURCE:
;     ORGANISM:  Homo sapiens
;     INDIVIDUAL ISOLATE:  Caucasian
;     TISSUE TYPE:  Blood
;   IMMEDIATE SOURCE:
;     CLONE:  Mfd14
;   POSITION IN GENOME:
;     CHROMOSOME/SEGMENT:  9
;   FEATURE:
;     NAME/KEY:  repeat_region
;     LOCATION:  82..128
;     OTHER INFORMATION:  /rpt_type= "tandem"
;     OTHER INFORMATION:  /rpt_family= "(dC-dA)n.(dG-dT)n"
;     OTHER INFORMATION:  /citation= ([2])
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  58..76
;     IDENTIFICATION METHOD:  experimental
;     OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     OTHER INFORMATION:  /standard_name= "PCR primer"
;     OTHER INFORMATION:  /citation= ([1])
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  complement (131..150)
;     IDENTIFICATION METHOD:  experimental
;     OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     OTHER INFORMATION:  /standard_name= "PCR primer"
;     OTHER INFORMATION:  /citation= ([1])
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  1..175
;     IDENTIFICATION METHOD:  experimental
;     OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     OTHER INFORMATION:  /standard_name= "Only one strand sequenced"
;   PUBLICATION INFORMATION:
;     AUTHORS:  Weber, J. L.
;     AUTHORS:  May, P. E.
;     TITLE:  Dinucleotide repeat polymorphism at the D9S43
;     TITLE:  locus
;     JOURNAL:  Nucleic Acids Res.
;     VOLUME:  18
;     PAGES:  2203-
;     DATE:  1990
;   PUBLICATION INFORMATION:
;     AUTHORS:  Weber, James L.
;     AUTHORS:  May, Paula E.
;     TITLE:  Abundant Class of Human DNA Polymorphisms
;     TITLE:  Which Can Be Typed Using the Polymerase Chain
;     TITLE:  Reaction
;     JOURNAL:  Am. J. Hum. Genet.
;     VOLUME:  44
;     PAGES:  388-396
;     DATE:  1989
US-08-222-177A-4

  Query Match             23.6%;  Score 26.2;  DB 1;  Length 175;
  Best Local Similarity   53.4%;  Pred. No. 1.8;
  Matches   55;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy       2 TGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGG 61
             ||| | | |  | |    | |||||| || |  || | | ||   |       | || |
Db      13 TGGTTTGCTGGTGTTTTGTTCAGGATTTTTGTGTATATGTTCATCAAGGATATTGTCCTG 72

Qy      62 AGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTG 104
           || || |        || || | |||||||  || |||  ||
Db      73 AGGATTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 115


RESULT 4
US-08-666-367B-4/c
; Sequence 4, Application US/08666367B
; Patent No. 5854042
;  GENERAL INFORMATION:
;    APPLICANT:  Shuichi TSUJI et al.
;    TITLE OF INVENTION:  NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;    TITLE OF INVENTION:  PRODUCING THE SAME
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
```

```
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Wordperfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/666,367B
;       FILING DATE:  August 19, 1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warren M. Cheek, Jr.
;       REGISTRATION NUMBER:  33,367
;       REFERENCE/DOCKET NUMBER:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-371-8850
;       TELEFAX:
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
        LENGTH:  1146
        TYPE:  nucleic acid
        STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  mouse
;     FEATURE:
;       OTHER INFORMATION:  1-1128 sialyltransferase in soluble
;       OTHER INFORMATION:  form
US-08-666-367B-4

  Query Match             23.6%;  Score 26.2;  DB 2;  Length 1146;
  Best Local Similarity   63.5%;  Pred. No. 3.5;
  Matches   40;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      19 TTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 78
           || | | ||  | |  |||  || |   | | ||||||||  ||| ||||||||   |||
Db     840 TTCCTGAATGATGTCCCATAGTTCCCATGGCATCTGGGGCTTGAGGATGTAAAAAGGCTG 781

Qy      79 GGT 81
           | |
Db     780 GCT 778


RESULT 5
US-09-143-438-4/c
; Sequence 4, Application US/09143438
; Patent No. 6218161
;  GENERAL INFORMATION:
;    APPLICANT:  Shuichi TSUJI et al.
;    TITLE OF INVENTION:  NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
;    TITLE OF INVENTION:  PRODUCING THE SAME
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/143,438
;      FILING DATE:  August 28, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/666,367
;      FILING DATE:  August 19, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1146
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;    FEATURE:
;      OTHER INFORMATION:  1-1128 sialyltransferase in soluble
;      OTHER INFORMATION:  form
US-09-143-438-4

  Query Match             23.6%;  Score 26.2;  DB 4;  Length 1146;
  Best Local Similarity   63.5%;  Pred. No. 3.5;
  Matches   40;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      19 TTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTG 78
           || | | ||  | |  |||  || |   | | ||||||||  ||| ||||||||   |||
Db     840 TTCCTGAATGATGTCCCATAGTTCCCATGGCATCTGGGGCTTGAGGATGTAAAAAGGCTG 781

Qy      79 GGT 81
           | |
Db     780 GCT 778


RESULT 6
US-08-961-527-34/c
; Sequence 34, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13104 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
US-08-961-527-34

  Query Match             23.6%;  Score 26.2;  DB 4;  Length 13104;
  Best Local Similarity   67.3%;  Pred. No. 8.6;
  Matches   37;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      10 TCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGT 64
           ||| ||| |||| || |||||   ||| |||| |||||  | |  |||  | ||
Db    9341 TCTATTGTCTTGGGGGGTTCTTGGGCATGTTTGTAGGGATGTTTAAGGCACGTGT 9287


RESULT 7
US-09-397-787-254/c
; Sequence 254, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: King, Gordon E.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
;  TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
;  FILE REFERENCE: 210121.466C2
;  CURRENT APPLICATION NUMBER: US/09/397,787
;  CURRENT FILING DATE: 1999-09-16
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 254
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(430)
;   OTHER INFORMATION: n = A,T,C or G
US-09-397-787-254

  Query Match             23.4%;  Score 26;  DB 4;  Length 430;
  Best Local Similarity   70.0%;  Pred. No. 2.9;
  Matches   35;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      62 AGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||||| |||  ||||   |||||  ||||| |||  ||||   |||| ||
Db     402 AGTATCTAAGGCTCCCAAGTCTCCATGTGTCCCTTTGTGGGGCCTCTGCT 353


RESULT 8
US-08-991-840A-1/c
; Sequence 1, Application US/08991840A
; Patent No. 6261570
;  GENERAL INFORMATION:
;    APPLICANT:  Michael D. Parker
;    APPLICANT:  Jonathan F. Smith
;    APPLICANT:  Bruce Crise
;    APPLICANT:  Mark Steve Oberste
;    APPLICANT:  Shannon Schmura
;    TITLE OF INVENTION:  Live Attenuated Virus Vaccines for Eastern Equine Encephalitis
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Charles H. Harris
;      STREET:  USA MRMC -  MRMC-JA
;      CITY:  FORT DETRICK, FREDERICK
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21702-5012
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5
;      SOFTWARE:  Microsoft Word 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,840A
;      FILING DATE:  December 16, 1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  Provisional Application 60/047162,
;      FILING DATE:  May 20, 1997
;      APPLICATION NUMBER:  Provisional Application 60/053,652
;      FILING DATE:  July 24, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles H. Harris
;      REGISTRATION NUMBER:  34,616
;      REFERENCE/DOCKET NUMBER:  003/058/SAP RIID 96-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 619-2065
;      TELEFAX:  (301) 619-5034
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11492 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
;    FEATURE:  OTHER INFORMATION:  N at all occurences is = unknown.
;    FEATURE:  OTHER INFORMATION: K at all
;    FEATURE:  occurences is = G or T
US-08-991-840A-1

  Query Match             22.9%;  Score 25.4;  DB 4;  Length 11492;
  Best Local Similarity   64.4%;  Pred. No. 16;
  Matches   38;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      12 TTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAA 70
           ||| || | | || | |||| ||| |||| |  ||     |||| | |||| || || |
Db    7692 TTTGGCTTGGGAGCACTCTTCTTCTTCTTCGGCGGCGGACCTGGCGGCGGATTAGGTGA 7634


RESULT 9
US-08-428-713-1/c
; Sequence 1, Application US/08428713
; Patent No. 5866541
;  GENERAL INFORMATION:
;    APPLICANT:  HOOK, Magnus
;    APPLICANT:  LINDBERG, Kjell Martin
;    APPLICANT:  LINDGREN, Per-Eric
;    APPLICANT:  SIGNAS, Lars Christer
;    TITLE OF INVENTION:  FIBRONECTIN BINDING PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,713
;      FILING DATE:  25-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,222
;      FILING DATE:  23-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa Stanek
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  012885-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  1371 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-428-713-1

  Query Match             22.3%;  Score 24.8;  DB 2;  Length 1371;
  Best Local Similarity   56.0%;  Pred. No. 12;
  Matches   47;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      13 TTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAA 72
           | || |   || |||| ||||    || ||  ||  ||||||| |  |  || | |  ||
Db     322 TCTGACCACCAAGATTATTTTCTCCCTCTGTTGGACCTTCTGGTGTTGATGTCTCTGGAA 263

Qy      73 CTCCTGGGTCTCTGTGTGTGCCTG 96
           ||   || |||    || |   ||
Db     262 CTGAAGGTTCTGACGGTATCAGTG 239


ULT 10
08-904-179-1/c
; Sequence 1, Application US/08904179
; Patent No. 6086895
;  GENERAL INFORMATION:
;    APPLICANT:  HOOK, Magnus
;    APPLICANT:  LINDBERG, Kjell Martin
;    APPLICANT:  LINDGREN, Per-Eric
;    APPLICANT:  SIGNAS, Lars Christer
;    TITLE OF INVENTION:  FIBRONECTIN BINDING PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,179
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,713
;      FILING DATE:  25-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa Stanek
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  012885-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1371 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-904-179-1

  Query Match             22.3%;  Score 24.8;  DB 3;  Length 1371;
  Best Local Similarity   56.0%;  Pred. No. 12;
  Matches   47;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      13 TTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAA 72
           | || |   || |||| ||||    || ||  ||  ||||||| |  |  || | |  ||
Db     322 TCTGACCACCAAGATTATTTTCTCCCTCTGTTGGACCTTCTGGTGTTGATGTCTCTGGAA 263

Qy      73 CTCCTGGGTCTCTGTGTGTGCCTG 96
           ||   || |||    || |   ||
Db     262 CTGAAGGTTCTGACGGTATCAGTG 239


RESULT 11
US-08-428-713-9/c
; Sequence 9, Application US/08428713
; Patent No. 5866541
;  GENERAL INFORMATION:
;    APPLICANT:  HOOK, Magnus
;    APPLICANT:  LINDBERG, Kjell Martin
;    APPLICANT:  LINDGREN, Per-Eric
;    APPLICANT:  SIGNAS, Lars Christer
;    TITLE OF INVENTION:  FIBRONECTIN BINDING PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,713
;      FILING DATE:  25-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,222
;      FILING DATE:  23-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa Stanek
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  012885-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1374 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1164
US-08-428-713-9

  Query Match             22.3%;  Score 24.8;  DB 2;  Length 1374;
  Best Local Similarity   56.0%;  Pred. No. 12;
  Matches   47;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      13 TTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAA 72
           | || |   || |||| ||||    || ||  ||  ||||||| |  |  || | |  ||
Db     322 TCTGACCACCAAGATTATTTTCTCCCTCTGTTGGACCTTCTGGTGTTGATGTCTCTGGAA 263

Qy      73 CTCCTGGGTCTCTGTGTGTGCCTG 96
           ||   || |||    || |   ||
Db     262 CTGAAGGTTCTGACGGTATCAGTG 239


RESULT 12
US-08-904-179-9/c
; Sequence 9, Application US/08904179
; Patent No. 6086895
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  HOOK, Magnus
;    APPLICANT:  LINDBERG, Kjell Martin
;    APPLICANT:  LINDGREN, Per-Eric
;    APPLICANT:  SIGNAS, Lars Christer
;    TITLE OF INVENTION:  FIBRONECTIN BINDING PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/904,179
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,713
;      FILING DATE:  25-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rea, Teresa Stanek
;      REGISTRATION NUMBER:  30,427
;      REFERENCE/DOCKET NUMBER:  012885-074
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1374 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1164
US-08-904-179-9

  Query Match             22.3%;  Score 24.8;  DB 3;  Length 1374;
  Best Local Similarity   56.0%;  Pred. No. 12;
  Matches   47;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

       13 TTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAA 72
          | || |   || |||| ||||    || ||  ||  ||||||| |  |  || | |  ||
Db    322 TCTGACCACCAAGATTATTTTCTCCCTCTGTTGGACCTTCTGGTGTTGATGTCTCTGGAA 263

Qy     73 CTCCTGGGTCTCTGTGTGTGCCTG 96
          ||   || |||    || |   ||
Db    262 CTGAAGGTTCTGACGGTATCAGTG 239


RESULT 13
US-09-556-877-136/c
; Sequence 136, Application US/09556877
; Patent No. 6432916
; GENERAL INFORMATION:
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Skeiky, Yasir
;  APPLICANT:  Fling, Steve
;  APPLICANT:  Maisonneuve, Jeff
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF CHLAMYDIAL INFECTION
;  FILE REFERENCE:  210121.469C5
;  CURRENT APPLICATION NUMBER: US/09/556,877
;  CURRENT FILING DATE:  2000-04-19
;  NUMBER OF SEQ ID NOS:  305
;  SOFTWARE:  FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 136
;   LENGTH: 882
;   TYPE: DNA
;   ORGANISM: Chlamydia
US-09-556-877-136

  Query Match             22.2%;  Score 24.6;  DB 4;  Length 882;
  Best Local Similarity   76.9%;  Pred. No. 12;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      8 GATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGG 46
          |||||||  ||||| || ||  || ||| |||||||  ||
Db    367 GATCTTTCACCTTGGAGCATGTTTCTTCTTCTTTGCCGG 329


RESULT 14
US-09-620-412C-136/c
; Sequence 136, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
;  APPLICANT: Steven P. Fling
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF CHLAMYDIAL INFECTION
;  FILE REFERENCE: 210121.469C7
;  CURRENT APPLICATION NUMBER: US/09/620,412C
;  CURRENT FILING DATE:  2000-07-20
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 136
;   LENGTH: 882
;   TYPE: DNA
;   ORGANISM: Chlamydia
US-09-620-412C-136

  Query Match             22.2%;  Score 24.6;  DB 4;  Length 882;
  Best Local Similarity   76.9%;  Pred. No. 12;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      8 GATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGG 46
          |||||||  ||||| || ||  || ||| |||||||  ||
Db    367 GATCTTTCACCTTGGAGCATGTTTCTTCTTCTTTGCCGG 329


RESULT 15
US-09-370-807-7
; Sequence 7, Application US/09370807
; Patent No. 6297034
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Falco, S. Carl
;  APPLICANT: Rafalski, J. Antoni
;  APPLICANT: Sakai, Hajime
;  TITLE OF INVENTION: N-End Rule Pathway Enzymes
;  FILE REFERENCE: BB-1199
;  CURRENT APPLICATION NUMBER: US/09/370,807
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: 60/096,225
;  EARLIER FILING DATE: August 12, 1998
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2407
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-370-807-7

  Query Match             22.2%;  Score 24.6;  DB 4;  Length 2407;
  Best Local Similarity   57.0%;  Pred. No. 17;
  Matches   45;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     30 TTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTCTGTGT 89
          |||| || |  ||||  || |||   | |   || |  ||| ||| |||| | |
```

```
Db    882 TTTTACAGCACTGCATTGAATTCAAAGACAACAGCAATTAAGACTGCTGGCTTTAAAACA 941

Qy     90 GTGCCTGAGTGGCTGCTCT 108
          | ||||| |  ||| | |
Db    942 TTTCCTGATTACCTGGTGT 960


Search completed: April 25, 2003, 02:29:59
Job time : 67 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 02:28:45 ; Search time 69 Seconds
(without alignments)
1750.468 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161
Perfect score: 111
Sequence: 1 atgggtggatcttttgcctt.....gcctgagtggctgctctact 111

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

1 number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	58.2	52.4	570	10	US-09-864-761-9118	Sequence 9118, Ap
c	2	57.6	51.9	684973	10	US-09-263-959-1	Sequence 1, Appli
	3	52.8	47.6	428	9	US-09-918-995-8153	Sequence 8153, Ap
	4	38.8	35.0	162	10	US-09-864-761-31393	Sequence 31393, A
	5	38.8	35.0	519	10	US-09-864-761-14866	Sequence 14866, A
c	6	29.2	26.3	369	10	US-09-954-456-922	Sequence 922, App
c	7	29.2	26.3	369	10	US-09-954-456-1538	Sequence 1538, Ap
	8	29.2	26.3	1314	9	US-10-125-540-609	Sequence 609, App
	9	29.2	26.3	1314	10	US-09-764-870-609	Sequence 609, App
	10	28	25.2	342	9	US-10-125-540-118	Sequence 118, App
	11	28	25.2	342	10	US-09-764-870-118	Sequence 118, App
	12	28	25.2	412	10	US-09-783-590-1450	Sequence 1450, Ap
c	13	27.2	24.5	504	9	US-10-092-154-1665	Sequence 1665, Ap
c	14	27.2	24.5	504	9	US-10-092-154-1666	Sequence 1666, Ap
c	15	27.2	24.5	504	9	US-10-092-154-1667	Sequence 1667, Ap
c	16	27.2	24.5	504	10	US-09-764-847-1665	Sequence 1665, Ap
c	17	27.2	24.5	504	10	US-09-764-847-1666	Sequence 1666, Ap
c	18	27.2	24.5	504	10	US-09-764-847-1667	Sequence 1667, Ap
	19	27	24.3	487	10	US-09-747-155-270	Sequence 270, App
	20	26.6	24.0	6372	10	US-09-880-107-3948	Sequence 3948, Ap
c	21	26.4	23.8	385	10	US-09-770-791-90	Sequence 90, Appl
c	22	26.4	23.8	1656	9	US-09-938-842A-966	Sequence 966, App
	23	26.4	23.8	170834	10	US-09-835-232-7	Sequence 7, Appli
c	24	26.2	23.6	26225	9	US-10-091-504-1276	Sequence 1276, Ap
c	25	26.2	23.6	26225	10	US-09-764-869-1276	Sequence 1276, Ap
c	26	26	23.4	430	10	US-09-876-889-254	Sequence 254, App
	27	26	23.4	1548	10	US-09-925-302-211	Sequence 211, App
	28	26	23.4	1620	10	US-09-954-456-325	Sequence 325, App
	29	26	23.4	5641	9	US-10-015-219-1733	Sequence 1733, Ap
	30	26	23.4	5709	9	US-10-015-219-1734	Sequence 1734, Ap
c	31	25.8	23.2	590	10	US-09-864-761-12014	Sequence 12014, A
c	32	25.8	23.2	1663	10	US-09-822-830A-348	Sequence 348, App
c	33	25.6	23.1	440	9	US-10-092-154-366	Sequence 366, App
c	34	25.6	23.1	440	10	US-09-764-847-366	Sequence 366, App
	35	25.6	23.1	442	10	US-09-880-107-1917	Sequence 1917, Ap
	36	25.6	23.1	1184	10	US-09-969-347-210	Sequence 210, App
	37	25.6	23.1	32169	9	US-10-092-154-1963	Sequence 1963, Ap
	38	25.6	23.1	32169	10	US-09-764-847-1963	Sequence 1963, Ap
	39	25.4	22.9	11617	9	US-09-860-670-265	Sequence 265, App
	40	25.2	22.7	390	10	US-09-770-791-11	Sequence 11, Appl
c	41	25.2	22.7	455	10	US-09-960-352-9392	Sequence 9392, Ap
	42	25.2	22.7	509	9	US-09-992-598-149	Sequence 149, App
	43	25.2	22.7	509	9	US-09-989-293A-149	Sequence 149, App
	44	25.2	22.7	509	9	US-09-989-735-149	Sequence 149, App
	45	25.2	22.7	509	9	US-09-990-444-149	Sequence 149, App

ALIGNMENTS

RESULT 1
US-09-864-761-9118/c
; Sequence 9118, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

```
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9118
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000053.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   4
    OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.4
US-09-864-761-9118

  Query Match             52.4%;  Score 58.2;  DB 10;  Length 570;
  Best Local Similarity   79.3%;  Pred. No. 1.2e-10;
  Matches   69;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           | ||   | | |  | |||| |||| | || |||| ||||||||||||| ||||||||||
Db     566 GAATGGATCTCCTGCCTTGCTGGGATTCCTTGGGCTGGAGTATGTAAAATTCCTGGGTCT 507

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           |||||||||||||||||| ||||| |
Db     506 TTGTGTGTGCCTGAGTGGCCGCTCTGC 480


RESULT 2
US-09-263-959-1/c
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
;  GENERAL INFORMATION:
;    APPLICANT:  Hood, Leroy E.
;    APPLICANT:  Rowen, Lee
;    APPLICANT:  Koop, Ben F.
;    TITLE OF INVENTION:  DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
     NUMBER OF SEQUENCES:  1279
     CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/263,959
;      FILING DATE:  05-MAR-1999
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  920010.426C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  684973 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-263-959-1

  Query Match             51.9%;  Score 57.6;  DB 10;  Length 684973;
  Best Local Similarity   78.4%;  Pred. No. 1.5e-09;
  Matches   69;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      24 GGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCT 83
           |||    ||| |  | |||| |||  | ||| ||| ||| | ||||||||||||||||||
Db 404732 GGACAGATTTCCTGCCTTGCTGGGGATCCTGAGGCTGGAATGTGTAAAACTCCTGGGTCT 404673

Qy      84 CTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           |||||||||||||||  |||||||| ||
Db 404672 CTGTGTGTGCCTGAGCAGCTGCTCTGCT 404645


RESULT 3
US-09-918-995-8153
; Sequence 8153, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8153
;   LENGTH: 428
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-8153

  Query Match             47.6%;  Score 52.8;  DB 9;  Length 428;
  Best Local Similarity   71.9%;  Pred. No. 8.2e-09;
  Matches   69;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      15 TGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACT 74
           ||||||||  ||       ||| ||||   |||  | ||  ||| ||||||||||||  ||
Db     272 TGCCTTGCCTGAGATGCCATCACCTTTCTTGGGGATCCTAAGGCTGGAGTATGTAAAGCT 331

Qy      75 CCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTAC 110
           .|||||||||||||| ||||||||||   | | ||| |
Db     332 CCTGGGTCTCTGTATGTGCCTGAGCACCGGTTCTTC 367


RESULT 4
US-09-864-761-31393
; Sequence 31393, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
```

```
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31393
;   LENGTH: 162
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC017089.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
;   OTHER INFORMATION: SWISSPROT HIT: P98161, EVALUE 3.00e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AI792950.1, EVALUE 6.00e-07
;   OTHER INFORMATION: NT HIT: AL163210.2, EVALUE 4.00e-04
US-09-864-761-31393

  Query Match             35.0%;  Score 38.8;  DB 10;  Length 162;
  Best Local Similarity   70.3%;  Pred. No. 0.00043;
  Matches   52;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      38 CTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGA 97
           |||| |  |||  ||  ||||| || ||| |||| |||  | || |   ||||||||||
Db      52 CTTTTCCTGGAAGTCCAGGGCCAGAATATCTAAAGCTCTGGAGTTTTCATGTGTGCCTGA 111

Qy      98 GTGGCTGCTCTACT 111
           |  | ||||||  ||
Db     112 GCAGATGCTCTGCT 125


RESULT 5
US-09-864-761-14866
; Sequence 14866, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14866
;   LENGTH: 519
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC017089.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
US-09-864-761-14866

  Query Match             35.0%;  Score 38.8;  DB 10;  Length 519;
  Best Local Similarity   70.3%;  Pred. No. 0.0006;
  Matches   52;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      38 CTTTGCAGGGACTTCTGGGGCCGGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGA 97
           |||| |  |||  ||  ||||| || ||| |||| |||  | || |   ||||||||||
Db     301 CTTTTCCTGGAAGTCCAGGGCCAGAATATCTAAAGCTCTGGAGTTTTCATGTGTGCCTGA 360

Qy      98 GTGGCTGCTCTACT 111
           |  | ||||||  ||
Db     361 GCAGATGCTCTGCT 374


RESULT 6
US-09-954-456-922/c
; Sequence 922, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
```

```
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
;  SEQ ID NO 922
;    LENGTH: 369
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-954-456-922

  Query Match             26.3%;  Score 29.2;  DB 10;  Length 369;
  Best Local Similarity   57.8%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          | || ||||  ||||   || ||  ||    || || ||| |||   || ||| |  |
Db    143 AGGGCTGGAAATTTTAGTTTTCAATATAAGCTTCCAGCTTAGCAATTACCTCTAGTCCAA 84

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 90
          ||  || |   | |||| | ||| | || |
Db     83 GACAATATTTGATTCCTAGTTCTGTTTGGG 54


RESULT 7
US-09-954-456-1538/c
; Sequence 1538, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
;  SEQ ID NO 1538
;    LENGTH: 369
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-954-456-1538

  Query Match             26.3%;  Score 29.2;  DB 10;  Length 369;
  Best Local Similarity   57.8%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          | || ||||  ||||   || ||  ||    || || ||| |||   || ||| |  |
Db    143 AGGGCTGGAAATTTTAGTTTTCAATATAAGCTTCCAGCTTAGCAATTACCTCTAGTCCAA 84

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTG 90
          ||  || |   | |||| | ||| | || |
Db     83 GACAATATTTGATTCCTAGTTCTGTTTGGG 54


RESULT 8
US-10-125-540-609
; Sequence 609, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14C1
;  CURRENT APPLICATION NUMBER: US/10/125,540
;  CURRENT FILING DATE:  2002-04-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 609
;    LENGTH: 1314
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-125-540-609

  Query Match             26.3%;  Score 29.2;  DB 9;  Length 1314;
  Best Local Similarity   56.1%;  Pred. No. 1.6;
  Matches   55;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      3 GGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGA 62
          |||  || |  | | | ||  |||    ||  | || |  || | |    |||  |  |
Db    982 GGGGTGAACAGTAGGCCTGGGGGAGAAATTGCCTTCCTGACAAGAAGAGTTGGAGATGAT 1041

Qy     63 GTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTG 100
          | ||  ||||| |||||| | |||  ||  ||  ||||
Db   1042 GAATACAAAACCCCTGGGACACTGCCTGGCCCATAGTG 1079


RESULT 9
US-09-764-870-609
; Sequence 609, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 609
;    LENGTH: 1314
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-764-870-609

  Query Match             26.3%;  Score 29.2;  DB 10;  Length 1314;
  Best Local Similarity   56.1%;  Pred. No. 1.6;
  Matches   55;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;
```

```
Qy      3 GGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGA 62
          ||| || | | | | ||  |||    ||  | || |  || | |    ||| | |
Db    982 GGGGTGAACAGTAGGCCTGGGGGAGAAATTGCCTTCCTGACAAGAAGAGTTGGAGATGAT 1041

Qy     63 GTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTG 100
          | ||  ||||| |||||| | |||  ||  ||  ||||
Db   1042 GAATACAAAACCCCTGGGACACTGCCTGGCCCATAGTG 1079


RESULT 10
US-10-125-540-118
; Sequence 118, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14C1
;  CURRENT APPLICATION NUMBER: US/10/125,540
;  CURRENT FILING DATE:  2002-04-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 118
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (134)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (286)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-125-540-118

  Query Match             25.2%;  Score 28;  DB 9;  Length 342;
  Best Local Similarity   55.0%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          | | |  || |  | | | ||  |||    ||  | || |  || | |    ||| | |
Db      9 ACGAGGTGAACAGTAGGCCTGGGGGAGAAATTGCCTTCCTGACAAGAAGAGTTGGAGATG 68

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTG 100
          | ||  ||||| |||||| | |||  ||  ||  ||||
       69 ATGAATACAAAACCCCTGGGACACTGCCTGGCCCATAGTG 108


RESULT 11
US-09-764-870-118
; Sequence 118, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ14
;  CURRENT APPLICATION NUMBER: US/09/764,870
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 646
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 118
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (134)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (286)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-870-118

  Query Match             25.2%;  Score 28;  DB 10;  Length 342;
  Best Local Similarity   55.0%;  Pred. No. 2.9;
  Matches   55;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 ATGGGTGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCG 60
          | | |  || |  | | | ||  |||    ||  | || |  || | |    ||| | |
Db      9 ACGAGGTGAACAGTAGGCCTGGGGGAGAAATTGCCTTCCTGACAAGAAGAGTTGGAGATG 68

Qy     61 GAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTG 100
          | ||  ||||| |||||| | |||  ||  ||  ||||
Db     69 ATGAATACAAAACCCCTGGGACACTGCCTGGCCCATAGTG 108


RESULT 12
US-09-783-590-1450
; Sequence 1450, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
;  APPLICANT: Dillon, Patrick J.
;  APPLICANT:  Haseltine, William A.
;  APPLICANT:  Li, Haodong
;  APPLICANT:  Rosen, Craig A.
;  APPLICANT:  Ruben, Steven M.
;  TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;  FILE REFERENCE: PO-16.2C1
;  CURRENT APPLICATION NUMBER: US/09/783,590
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: 08/420,856
;  PRIOR FILING DATE: 1995-04-12
;  PRIOR APPLICATION NUMBER: 08/346,731
;  PRIOR FILING DATE: 1994-11-21
;  NUMBER OF SEQ ID NOS: 12485
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1450
;   LENGTH: 412
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (5)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (189)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (190)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (231)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (265)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (266)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (268)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (269)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (274)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (275)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (339)
```

```
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (340)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (370)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (399)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (402)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (403)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (406)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-1450

  Query Match             25.2%;  Score 28;  DB 10;  Length 412;
  Best Local Similarity   67.3%;  Pred. No. 3;
  Matches   37;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      57 GCCGGAGTATGTAAAACTCCTGGGTCTCTGTGTGTGCCTGAGTGGCTGCTCTACT 111
           ||| |||||| | |  ||  ||  ||| ||     |||| ||||||||||||  |
Db     213 GCCAGAGTATCTGAGGCTNTTGAATCTGTGCACAGGCCTCAGTGGCTGCTCTNNT 267


RESULT 13
US-10-092-154-1665/c
; Sequence 1665, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1665
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1665

  Query Match             24.5%;  Score 27.2;  DB 9;  Length 504;
  Best Local Similarity   56.8%;  Pred. No. 6.1;
  Matches   50;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       6 TGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTA 65
           | | |  |||| | | |||   || ||||| |     ||| | |  ||   | |  ||
Db     321 TTGGTTGTTTGTCCTTCAGTCCTCCTTTTCTTAAACACAGAGCCAGCTCCTGGCATAGGC 262

Qy      66 TGTAAAACTCCTGGGTCTCTGTGTGTGC 93
             | ||   | || |||| | ||||||||
Db     261 AGGAACTATGCTAGGTCCCAGTGTGTGC 234


RESULT 14
US-10-092-154-1666/c
; Sequence 1666, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1666
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1666

  Query Match             24.5%;  Score 27.2;  DB 9;  Length 504;
  Best Local Similarity   56.8%;  Pred. No. 6.1;
  Matches   50;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       6 TGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTA 65
           | | |  |||| | | |||   || ||||| |     ||| | |  ||   | |  ||
Db     321 TTGGTTGTTTGTCCTTCAGTCCTCCTTTTCTTAAACACAGAGCCAGCTCCTGGCATAGGC 262

Qy      66 TGTAAAACTCCTGGGTCTCTGTGTGTGC 93
             | ||   | || |||| | ||||||||
Db     261 AGGAACTATGCTAGGTCCCAGTGTGTGC 234


RESULT 15
US-10-092-154-1667/c
; Sequence 1667, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009C1
;  CURRENT APPLICATION NUMBER: US/10/092,154
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2003
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1667
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1667

  Query Match             24.5%;  Score 27.2;  DB 9;  Length 504;
  Best Local Similarity   56.8%;  Pred. No. 6.1;
  Matches   50;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       6 TGGATCTTTTGCCTTGCAGGATTCTTTTTCATCTTTGCAGGGACTTCTGGGGCCGGAGTA 65
           | | |  |||| | | |||   || ||||| |     ||| | |  ||   | |  ||
Db     321 TTGGTTGTTTGTCCTTCAGTCCTCCTTTTCTTAAACACAGAGCCAGCTCCTGGCATAGGC 262

Qy      66 TGTAAAACTCCTGGGTCTCTGTGTGTGC 93
             | ||   | || |||| | ||||||||
Db     261 AGGAACTATGCTAGGTCCCAGTGTGTGC 234


Search completed: April 25, 2003, 03:07:48
Job time : 228 secs
```